

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
21 June 2001 (21.06.2001)

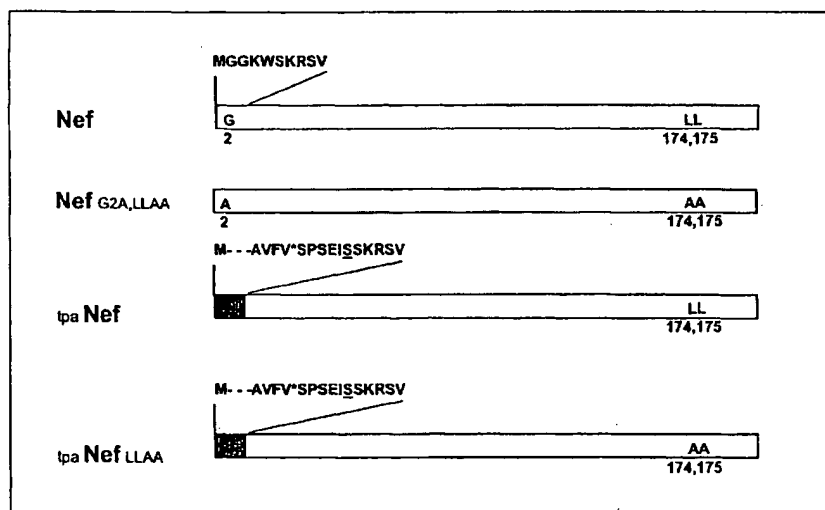
PCT

(10) International Publication Number
WO 01/43693 A2

- (51) International Patent Classification⁷: **A61K** [US/US]; 126 East Lincoln Avenue, Rahway, NJ 07065-0907 (US).
- (21) International Application Number: PCT/US00/34162
- (22) International Filing Date:
15 December 2000 (15.12.2000)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:
60/172,442 17 December 1999 (17.12.1999) US
- (71) Applicant (for all designated States except US): **MERCK & CO., INC.** [US/US]; 126 East Lincoln Avenue, Rahway, NJ 07065-0907 (US).
- (72) Inventors; and
- (75) Inventors/Applicants (for US only): **SHIVER, John, W.** [US/US]; 126 East Lincoln Avenue, Rahway, NJ 07065-0907 (US). **LIANG, Xiaoping** [CA/US]; 126 East Lincoln Avenue, Rahway, NJ 07065-0907 (US). **FU, Tong-Ming**
- (81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.
- (84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).
- Published:
— Without international search report and to be republished upon receipt of that report.

[Continued on next page]

(54) Title: POLYNUCLEOTIDE VACCINES EXPRESSING CODON OPTIMIZED HIV-1 NEF AND MODIFIED HIV-1 NEF



(57) Abstract: Pharmaceutical compositions which comprise HIV Nef DNA vaccines are disclosed, along with the production and use of these DNA vaccines. The nef-based DNA vaccines of the invention are administered directly introduced into living vertebrate tissue, preferably humans, and express the HIV Nef protein or biologically relevant portions thereof, inducing a cellular immune response which specifically recognizes human immunodeficiency virus-1 (HIV-1). The DNA molecules which comprise the open reading frame of these DNA vaccines are synthetic DNA molecules encoding codon optimized HIV-1 Nef and derivatives of optimized HIV-1 Nef, including nef modifications comprising amino terminal leader peptides, removal of the amino terminal myristylation site, and/or modification of the Nef dileucine motif. These modifications may effect wild type characteristics of Nef, such as myristylation and down regulation of host CD4.

WO 01/43693 A2



For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

TITLE OF THE INVENTION

5 POLYNUCLEOTIDE VACCINES EXPRESSING CODON OPTIMIZED HIV-1
NEF AND MODIFIED HIV-1 NEF

CROSS-REFERENCE TO RELATED APPLICATIONS

10 This application claims the benefit, under 35 U.S.C. §119(e), of U.S.
provisional application 60/172,442, filed December 17, 1999.

STATEMENT REGARDING FEDERALLY-SPONSORED R&D

Not Applicable

15

REFERENCE TO MICROFICHE APPENDIX

Not Applicable

FIELD OF THE INVENTION

20 The present invention relates to HIV Nef polynucleotide pharmaceutical
products, as well as the production and use thereof which, when directly introduced
into living vertebrate tissue, preferably a mammalian host such as a human or a
non-human mammal of commercial or domestic veterinary importance, express the
HIV Nef protein or biologically relevant portions thereof within the animal, inducing
25 a cellular immune response which specifically recognizes human immunodeficiency
virus-1 (HIV-1). The polynucleotides of the present invention are synthetic DNA
molecules encoding codon optimized HIV-1 Nef and derivatives of optimized HIV-1
Nef, including nef mutants which effect wild type characteristics of Nef, such as
myristylation and down regulation of host CD4. The polynucleotide vaccines of the
30 present invention should offer a prophylactic advantage to previously uninfected
individuals and/or provide a therapeutic effect by reducing viral load levels within an
infected individual, thus prolonging the asymptomatic phase of HIV-1 infection.

BACKGROUND OF THE INVENTION

Human Immunodeficiency Virus-1 (HIV-1) is the etiological agent of acquired human immune deficiency syndrome (AIDS) and related disorders. HIV-1 is an RNA virus of the Retroviridae family and exhibits the 5' LTR-*gag-pol-env*-LTR 3' organization of all retroviruses. The integrated form of HIV-1, known as the provirus, is approximately 9.8 Kb in length. Each end of the viral genome contains flanking sequences known as long terminal repeats (LTRs). The HIV genes encode at least nine proteins and are divided into three classes; the major structural proteins (Gag, Pol, and Env), the regulatory proteins (Tat and Rev); and the accessory proteins (Vpu, Vpr, Vif and Nef).

The *gag* gene encodes a 55-kilodalton (kDa) precursor protein (p55) which is expressed from the unspliced viral mRNA and is proteolytically processed by the HIV protease, a product of the *pol* gene. The mature p55 protein products are p17 (matrix), p24 (capsid), p9 (nucleocapsid) and p6.

The *pol* gene encodes proteins necessary for virus replication; a reverse transcriptase, a protease, integrase and RNase H. These viral proteins are expressed as a Gag-Pol fusion protein, a 160 kDa precursor protein which is generated via a ribosomal frame shifting. The viral encoded protease proteolytically cleaves the Pol polypeptide away from the Gag-Pol fusion and further cleaves the Pol polypeptide to the mature proteins which provide protease (Pro, P10), reverse transcriptase (RT, P50), integrase (IN, p31) and RNase H (RNase, p15) activities.

The *nef* gene encodes an early accessory HIV protein (Nef) which has been shown to possess several activities such as down regulating CD4 expression, disturbing T-cell activation and stimulating HIV infectivity.

The *env* gene encodes the viral envelope glycoprotein that is translated as a 160-kilodalton (kDa) precursor (gp160) and then cleaved by a cellular protease to yield the external 120-kDa envelope glycoprotein (gp120) and the transmembrane 41-kDa envelope glycoprotein (gp41). Gp120 and gp41 remain associated and are displayed on the viral particles and the surface of HIV-infected cells.

The *tat* gene encodes a long form and a short form of the Tat protein, a RNA binding protein which is a transcriptional transactivator essential for HIV-1 replication.

The *rev* gene encodes the 13 kDa Rev protein, a RNA binding protein. The Rev protein binds to a region of the viral RNA termed the Rev response element

(RRE). The Rev protein is promotes transfer of unspliced viral RNA from the nucleus to the cytoplasm. The Rev protein is required for HIV late gene expression and in turn, HIV replication.

Gp120 binds to the CD4/chemokine receptor present on the surface of helper
5 T-lymphocytes, macrophages and other target cells in addition to other co-receptor molecules. X4 (macrophage tropic) virus show tropism for CD4/CXCR4 complexes while a R5 (T-cell line tropic) virus interacts with a CD4/CCR5 receptor complex. After gp120 binds to CD4, gp41 mediates the fusion event responsible for virus entry. The virus fuses with and enters the target cell, followed by reverse transcription of its
10 single stranded RNA genome into the double-stranded DNA via a RNA dependent DNA polymerase. The viral DNA, known as provirus, enters the cell nucleus, where the viral DNA directs the production of new viral RNA within the nucleus, expression of early and late HIV viral proteins, and subsequently the production and cellular release of new virus particles. Recent advances in the ability to detect viral load
15 within the host shows that the primary infection results in an extremely high generation and tissue distribution of the virus, followed by a steady state level of virus (albeit through a continual viral production and turnover during this phase), leading ultimately to another burst of virus load which leads to the onset of clinical AIDS. Productively infected cells have a half life of several days, whereas chronically or
20 latently infected cells have a 3-week half life, followed by non-productively infected cells which have a long half life (over 100 days) but do not significantly contribute to day to day viral loads seen throughout the course of disease.

Destruction of CD4 helper T lymphocytes, which are critical to immune defense, is a major cause of the progressive immune dysfunction that is the hallmark
25 of HIV infection. The loss of CD4 T-cells seriously impairs the body's ability to fight most invaders, but it has a particularly severe impact on the defenses against viruses, fungi, parasites and certain bacteria, including mycobacteria.

Effective treatment regimens for HIV-1 infected individuals have become available recently. However, these drugs will not have a significant impact on the
30 disease in many parts of the world and they will have a minimal impact in halting the spread of infection within the human population. As is true of many other infectious diseases, a significant epidemiologic impact on the spread of HIV-1 infection will only occur subsequent to the development and introduction of an effective vaccine. There are a number of factors that have contributed to the lack of successful vaccine

development to date. As noted above, it is now apparent that in a chronically infected person there exists constant virus production in spite of the presence of anti-HIV-1 humoral and cellular immune responses and destruction of virally infected cells. As in the case of other infectious diseases, the outcome of disease is the result of a balance between the kinetics and the magnitude of the immune response and the pathogen replicative rate and accessibility to the immune response. Pre-existing immunity may be more successful with an acute infection than an evolving immune response can be with an established infection. A second factor is the considerable genetic variability of the virus. Although anti-HIV-1 antibodies exist that can neutralize HIV-1 infectivity in cell culture, these antibodies are generally virus isolate-specific in their activity. It has proven impossible to define serological groupings of HIV-1 using traditional methods. Rather, the virus seems to define a serological "continuum" so that individual neutralizing antibody responses, at best, are effective against only a handful of viral variants. Given this latter observation, it would be useful to identify immunogens and related delivery technologies that are likely to elicit anti-HIV-1 cellular immune responses. It is known that in order to generate CTL responses antigen must be synthesized within or introduced into cells, subsequently processed into small peptides by the proteasome complex, and translocated into the endoplasmic reticulum/Golgi complex secretory pathway for eventual association with major histocompatibility complex (MHC) class I proteins. CD8⁺ T lymphocytes recognize antigen in association with class I MHC via the T cell receptor (TCR) and the CD8 cell surface protein. Activation of naive CD8⁺ T cells into activated effector or memory cells generally requires both TCR engagement of antigen as described above as well as engagement of costimulatory proteins. Optimal induction of CTL responses usually requires "help" in the form of cytokines from CD4⁺ T lymphocytes which recognize antigen associated with MHC class II molecules via TCR and CD4 engagement.

As introduced above, the *nef* gene encodes an early accessory HIV protein (Nef) which has been shown to possess several activities such as down regulating CD4 expression, disturbing T-cell activation and stimulating HIV infectivity. Zazopoulos and Haseltine (1992, *Proc. Natl. Acad. Sci.* 89: 6634-6638) disclose mutations to the HIV-1 *nef* gene which effect the rate of virus replication. The authors show that the *nef* open reading frame mutated to encode Ala-2 in place of Gly-2 inhibits myristolation of the protein and results in delayed viral replication rates

in Jurkat cells and PBMCs.

Kaminchik et al. (1991, *J. Virology* 65(2): 583-588) disclose an amino-terminal nef open reading frame mutated to encode Met-Ala-Ala in place of Met-Gly-Gly. The authors show that this mutant is deficient in myristolation.

5 Saksela et al. (1995, *EMBO J.* 14(3): 484-491) and Lee et al. (1995, *EMBO J.* 14(20): 5006-5015) show the importance of a proline rich motif in HIV-1 Nef which mediates binding to a SH3 domain of the Hck protein. The authors conclude that this motif is important in the enhancement of viral replication but not down-regulation of CD4 expression.

10 Calarota et al. (1998, *The Lancet* 351: 1320-1325) present human clinical data concerning immunization of three HIV infected individuals with a DNA plasmid expressing wild type Nef. The authors conclude that immunization with a Nef encoding DNA plasmid induced a cellular immune response in the three individuals. However, two of the three patients were on alternative therapies during the study, and
15 the authors conclude that the CTL response was most likely a boost to a pre-existing CTL response. In addition, the viral load increased substantially in two of the three patients during the course of the study.

Tobery et al. (1997, *J. Exp. Med.* 185(5): 909-920) constructed two ubiquitin-nef (Ub-nef) fusion constructs, one which encoded the Nef initiating methionine and
20 the other with an Arg residue at the amino terminus of the Nef open reading frame. The authors state that vaccinia- or plasmid-based immunization of mice with a Ub-nef construct containing an Arg residue at the amino terminus induces a Nef-specific CTL response. The authors suggest the expressed fusion protein is more efficiently presented to the MHC class I antigen presentation pathway, resulting in an improved
25 cellular immune response.

Kim et al. (1997, *J. Immunol.* 158(2): 816-826) disclose that co-administration of a plasmid DNA construct expressing IL-12 with a plasmid construct expressing Nef results in an improved cellular immune response in mice when compared to inoculation with the Nef construct alone. The authors reported a reduction in the
30 humoral response from the Nef / IL-12 co-administration as compared to administration of the plasmid construct expressing Nef alone.

Moynier et al. (1998, *Vaccine* 16(16): 1523-1530) show varying humoral responses in mice immunized with a DNA plasmid encoding Nef, depending upon the presence of absence of Freund's adjuvant. No data is disclosed regarding a cellular

immune response in mice vaccinated with the aforementioned DNA construct alone.

Hanna et al. (1998, *Cell* 95:163-175) suggest that wild type Nef may play a critical role in AIDS pathogenicity.

- It would be of great import in the battle against AIDS to produce a
- 5 prophylactic- and/or therapeutic-based HIV vaccine which generates a strong cellular immune response against an HIV infection. The present invention addresses and meets this needs by disclosing a class of DNA vaccines based on host delivery and expression of the early HIV gene, *nef*.

10 SUMMARY OF THE INVENTION

- The present invention relates to synthetic DNA molecules (also referred to herein as "polynucleotides") and associated DNA vaccines (also referred to herein as "polynucleotide vaccines") which elicit CTL responses upon administration to the host, such as a mammalian host and including primates and especially humans, as
- 15 well as non-human mammals of commercial or domestic veterinary importance. The CTL-directed vaccines of the present invention should lower transmission rate to previously uninfected individuals and/or reduce levels of the viral loads within an infected individual, so as to prolong the asymptomatic phase of HIV-1 infection. In particular, the present invention relates to DNA vaccines which encode various forms
- 20 of HIV-1 Nef, wherein administration, intracellular delivery and expression of the HIV-1 *nef* gene of interest elicits a host CTL and Th response. The preferred synthetic DNA molecules of the present invention encode codon optimized versions of wild type HIV-1 Nef, codon optimized versions of HIV-1 Nef fusion proteins, and codon optimized versions of HIV-1 Nef derivatives, including but not limited to *nef*
- 25 modifications involving introduction of an amino-terminal leader sequence, removal of an amino-terminal myristylation site and/or introduction of dileucine motif mutations. The Nef-based fusion and modified proteins disclosed within this specification may possess altered trafficking and/or host cell function while retaining the ability to be properly presented to the host MHC I complex and in turn elicit a
- 30 host CTL and Th response.

A particular embodiment of the present invention relates to a DNA molecule encoding HIV-1 Nef from the HIV-1 jfrr isolate wherein the codons are optimized for expression in a mammalian system such as a human. The DNA molecule which encodes this protein is disclosed herein as SEQ ID NO:1, while the expressed open

reading frame is disclosed herein as SEQ ID NO:2.

In another embodiment of the present invention, a codon optimized DNA molecule encoding a protein containing the human plasminogen activator (tpa) leader peptide fused with the NH₂-terminus of the HIV-1 Nef polypeptide. The DNA molecule which encodes this protein is disclosed herein as SEQ ID NO:3, while the expressed open reading frame is disclosed herein as SEQ ID NO:4.

In an additional embodiment, the present invention relates to a DNA molecule encoding optimized HIV-1 Nef wherein the open reading frame codes for modifications at the amino terminal myristylation site (Gly-2 to Ala-2) and substitution of the Leu-174-Leu-175 dileucine motif to Ala-174-Ala-175, herein described as opt nef (G2A,LLAA). The DNA molecule which encodes this protein is disclosed herein as SEQ ID NO:5, while the expressed open reading frame is disclosed herein as SEQ ID NO:6.

Another additional embodiment of the present invention relates to a DNA molecule encoding optimized HIV-1 Nef wherein the amino terminal myristylation site and dileucine motif have been deleted, as well as comprising a tPA leader peptide. This DNA molecule, opt tpanef (LLAA), comprises an open reading frame which encodes a Nef protein containing a tPA leader sequence fused to amino acid residue 6-216 of HIV-1 Nef (jf1), wherein Leu-174 and Leu-175 are substituted with Ala-174 and Ala-175, herein referred to as opt tpanef (LLAA) is disclosed herein as SEQ ID NO:7, while the expressed open reading frame is disclosed herein as SEQ ID NO:8.

The present invention also relates to non-codon optimized versions of DNA molecules and associated DNA vaccines which encode the various wild type and modified forms of the HIV Nef protein disclosed herein. Partial or fully codon optimized DNA vaccine expression vector constructs are preferred, but it is within the scope of the present invention to utilize "non-codon optimized" versions of the constructs disclosed herein, especially modified versions of HIV Nef which are shown to promote a substantial cellular immune response subsequent to host administration.

The DNA backbone of the DNA vaccines of the present invention are preferably DNA plasmid expression vectors. DNA plasmid expression vectors utilized in the present invention include but are not limited to constructs which comprise the cytomegalovirus promoter with the intron A sequence (CMV-intA) and a bovine growth hormone transcription termination sequence. In addition, the DNA plasmid vectors of the present invention preferably comprise an antibiotic resistance

marker, including but not limited to an ampicillin resistance gene, a neomycin resistance gene or any other pharmaceutically acceptable antibiotic resistance marker. In addition, an appropriate polylinker cloning site and a prokaryotic origin of replication sequence are also preferred. Specific DNA vectors of the present invention include but are not limited to V1, VIJ (SEQ ID NO:14), V1Jneo (SEQ ID NO:15), V1Jns (Figure 1A, SEQ ID NO:16), V1R (SEQ ID NO:26), and any of the aforementioned vectors wherein a nucleotide sequence encoding a leader peptide, preferably the human tPA leader, is fused directly downstream of the CMV-intA promoter, including but not limited to V1Jns-tpa, as shown in Figure 1B and SEQ ID NO:19.

The present invention especially relates to a DNA vaccine and a pharmaceutically active vaccine composition which contains this DNA vaccine, and the use as a prophylactic and/or therapeutic vaccine for host immunization, preferably human host immunization, against an HIV infection or to combat an existing HIV condition. These DNA vaccines are represented by codon optimized DNA molecules encoding HIV-1 Nef of biologically active Nef modifications or Nef-containing fusion proteins which are ligated within an appropriate DNA plasmid vector, with or without a nucleotide sequence encoding a functional leader peptide. DNA vaccines of the present invention relate in part to codon optimized DNA molecules encoding HIV-1 Nef of biologically active Nef modifications or Nef-containing fusion proteins ligated in DNA vectors V1, VIJ (SEQ ID NO:14), V1Jneo (SEQ ID NO:15), V1Jns (Figure 1A, SEQ ID NO:16), V1R (SEQ ID NO:26), or any of the aforementioned vectors wherein a nucleotide sequence encoding a leader peptide, preferably the human tPA leader, is fused directly downstream of the CMV-intA promoter, including but not limited to V1Jns-tpa, as shown in Figure 1B and SEQ ID NO:19. Especially preferred DNA vaccines of the present invention include codon optimized DNA vaccine constructs V1Jns/nef, V1Jns/tpanef, V1Jns/tpanef(LLAA) and V1Jns/(G2A,LLAA), as exemplified in Example Section 2.

The present invention also relates to HIV Nef polynucleotide pharmaceutical products, as well as the production and use thereof, wherein the DNA vaccines are formulated with an adjuvant or adjuvants which may increase immunogenicity of the DNA polynucleotide vaccines of the present invention, namely by increasing a humoral response to inoculation. A preferred adjuvant is an aluminum phosphate-based adjuvant or a calcium phosphate based adjuvant,

with an aluminum phosphate adjuvant being especially preferred. Another preferred adjuvant is a non-ionic block copolymer, preferably comprising the blocks of polyoxyethylene (POE) and polyoxypropylene (POP) such as a POE-POP-POE block copolymer. These adjuvanted forms comprising the DNA vaccines disclosed herein are useful in increasing humoral responses to DNA vaccination without imparting a negative effect on an appropriate cellular immune response.

As used herein, a DNA vaccine or DNA polynucleotide vaccine or polynucleotide vaccine is a DNA molecule (i.e., "nucleic acid", "polynucleotide") which contains essential regulatory elements such that upon introduction into a living, vertebrate cell, it is able to direct the cellular machinery to produce translation products encoded by the respective nef genes of the present invention.

BRIEF DESCRIPTION OF THE FIGURES

Figure 1A-B show a schematic representation of DNA vaccine expression vectors V1Jns (A) and V1Jns/tpa utilized for HIV-1 nef and HIV-1 modified nef constructs.

Figure 2A-B show a nucleotide sequence comparison between wild type nef(jrfl) and codon optimized nef. The wild type nef gene from the jrfl isolate consists of 648 nucleotides capable of encoding a 216 amino acid polypeptide. WT, wild type sequence (SEQ ID NO:9); opt, codon-optimized sequence (contained within SEQ ID NO:1). The Nef amino acid sequence is shown in one-letter code (SEQ ID NO:2).

Figure 3A-C show nucleotide sequences at junctions between nef coding sequence and plasmid backbone of nef expression vectors V1Jns/nef (Figure 3A), V1Jns/nef(G2A,LLAA) (Figure 3B), V1Jns/tpanef (Figure 3C) and V1Jns/tpanef(LLAA) (Figure 3C, also). 5' and 3' flanking sequences of codon optimized nef or codon optimized nef mutant genes are indicated by bold/italic letters; nef and nef mutant coding sequences are indicated by plain letters. Also indicated (as underlined) are the restriction endonuclease sites involved in construction of respective nef expression vectors. V1Jns/tpanef and V1Jns/tpanef(LLAA) have identical sequences at the junctions.

Figure 4 shows a schematic presentation of nef and nef derivatives. Amino acid residues involved in Nef derivatives are presented. Glycine 2 and Leucine 174

and 175 are the sites involved in myristylation and dileucine motif, respectively. For both versions of the tpanef fusion genes, the putative leader peptide cleavage sites are indicated with "*", and a exogenous serine residue introduced during the construction of the mutants is underlined.

5 Figure 5 shows Western blot analysis of nef and modified nef proteins expressed in transfected 293 cells. 293 cells grown in 100 mm culture dish were transfected with respective codon optimized nef constructs. Sixty hours post transfection, supernatant and cells were collected separately and separated on 10% SDS-PAGE under reducing conditions. The proteins were transferred into a PVDF
10 membrane and probed with a mixture of Gag mAb and Nef mAbs, both at 1:2000 dilution. The protein signals were detected with ECL. (A) cells transfected with VIJns/gag only; (B) cells transfected with VIJns/gag and VIJns/nef; (C) cells transfected with VIJns/gag and VIJns/nef(G2A, LLAA); (D) cells transfected with VIJns/gag and VIJns/tpanef; (E) cells transfected with VIJns/gag and
15 VIJns/tpanef(LLAA). The low case letter c and m represent medium and cellular fractions, respectively. M.W. = molecular weight marker.

Figure 6 shows an Elispot assay of cell-mediated responses to Nef peptides. Three strains of mice, Balb/c, C57BL/6 and C3H, were immunized with 50 mcg of VIJns/nef (codon optimized) and boosted twice with a two-week interval. Two
20 weeks following the final immunization, splenocytes were isolated and tested in an Elispot assay against respective Nef peptide pools. As a control, splenocytes were from non-immunized naive mice were tested in parallel. Nef peptide pool A consists of all 21 Nef peptides; Nef peptide pool B consists of 11 non-overlapping peptide started from residue 1; Nef peptide pool C consists of 10 non-overlapping peptides
25 started from residue 11. SFC, INF-gamma secreting spot-forming cells.

Figure 7A-C show Nef-specific CD8 and CD4 epitope mapping. The immunization regime is as per Figure 6. Mouse splenocytes were isolated and fractionated into CD8⁺ and CD8⁻ cells using Miltenyi's magnetic cell separator. The resultant CD8⁺ and CD8⁻ cells were then tested in an Elispot assay against individual
30 Nef peptides. SFC, INF-gamma secreting spot-forming cells. The mice strains tested are Balb/c mice (Figure 7A), C57BL/6 mice (Figure 7B), and C3H mice (Figure 7C).

Figure 8A-C show identification of a Nef CTL epitope. Splenocytes from nef immunized C57BL/6 mice were stimulated *in vitro* with peptide-pulsed, irradiated naïve splenocytes for 7 days. Following the *in vitro* stimulation, cells were harvested

and tested in a standard ^{51}Cr -releasing assay using peptide pulsed EL-4 cells as targets. Open symbol, specific killings of EL-4 cells without peptide; solid symbol, specific killing of EL-4 cells with peptide. Panel A - peptide Nef 51-70; Panel B - peptide Nef 60-68, Panel C - peptide Nef 58-70.

- 5 Figure 9A-B shows a comparison of the immunogenicity of codon optimized DNA vaccine vectors expressing Nef and modified forms of Nef C57BL/6 mice, five per group, were immunized with 100 mcg of the indicated nef constructs. Fourteen days following immunization, splenocytes were collected and tested against the Nef CD8 (aa58-66) and CD4 (aa81-100) peptides. Identical immunization regimens were used for both experiments. In experiment 1 (Panel A), three codon optimized nef constructs were tested, namely, VIJns/nef, VIJns/tpanef(LLAA) and VIJns/nef(G2A,LLAA), whereas in experiment 2 (Panel B) all four codon optimized nef constructs were tested. The data represent means plus standard deviation of 5 mice per group.

15

DETAILED DESCRIPTION OF THE INVENTION

- The present invention relates to synthetic DNA molecules (also referred to herein as "nucleic acid" molecules or "polynucleotides") and associated DNA vector vaccines (also referred to herein as "polynucleotide vaccines") which elicit CTL and humoral responses upon administration to the host, including primates and especially humans. In particular, the present invention relates to DNA vector vaccines which encode various forms of HIV-1 Nef, wherein administration, intracellular delivery and expression of the HIV-1 nef gene of interest elicits a host CTL and Th response. The synthetic DNA molecules of the present invention encode codon optimized versions of wild type HIV-1 Nef, codon optimized versions of HIV-1 Nef fusion proteins, and codon optimized versions of HIV-1 Nef derivatives, including but not limited to *nef* modifications involving introduction of an amino-terminal leader sequence, removal of an amino-terminal myristylation site and/or introduction of dileucine motif mutations. In some instances the Nef-based fusion and modified proteins disclosed within this specification possess altered trafficking and/or host cell function while retaining the ability to be properly presented to the host MHC I complex. Those skilled in the art will recognize that the use of nef genes from HIV-2 strains which express Nef proteins having analogous function to HIV-1 Nef would be expected to generate immune responses analogous to those described herein for

30

HIV-1 constructs.

In order to generate a CTL response, the immunogen must be synthesized within (MHC I presentation) or introduced into cells (MHC II presentation). For intracellular synthesized immunogens, the protein is expressed and then processed
 5 into small peptides by the proteasome complex, and translocated into the endoplasmic reticulum/Golgi complex secretory pathway for eventual association with major histocompatibility complex (MHC) class I proteins. CD8⁺ T lymphocytes recognize antigen in association with class I MHC via the T cell receptor (TCR). Activation of naive CD8⁺ T cells into activated effector or memory cells generally requires both
 10 TCR engagement of antigen as described above as well as engagement of co-stimulatory proteins. Optimal induction of CTL responses usually requires "help" in the form of cytokines from CD4⁺ T lymphocytes which recognize antigen associated with MHC class II molecules via TCR.

The HIV-1 genome employs predominantly uncommon codons compared to
 15 highly expressed human genes. Therefore, the nef open reading frame has been synthetically manipulated using optimal codons for human expression. As noted above, a preferred embodiment of the present invention relates to DNA molecules which comprise a HIV-1 nef open reading frame, whether encoding full length nef or a modification or fusion as described herein, wherein the codon usage has been
 20 optimized for expression in a mammal, especially a human.

In a particular embodiment of the present invention, a DNA molecule encoding HIV-1 Nef from the HIV-1 jf1 isolate wherein the codons are optimized for expression in a mammalian system such as a human. The nucleotide sequence of the codon optimized version of HIV-1 jf1 nef gene is disclosed herein as SEQ ID NO:1,
 25 as shown herein:

```
GATCTGCCAC CATGGGCGGC AAGTGGTCCA AGAGGTCCGT GCCCGGCTGG TCCACCGTGA
GGGAGAGGAT GAGGAGGGCC GAGCCCGCCG CCGACAGGGT GAGGAGGACC GAGCCCGCCG
CCGTGGGCGT GGGCGCCGTG TCCAGGGACC TGGAGAAGCA CGGCGCCATC ACCTCTCCA
ACACCGCCGC CACCAACGCC GACTGCGCCT GGCTGGAGGC CCAGGAGGAC GAGGAGGTGG
30 GCTTCCCCGT GAGGCCCCAG GTGCCCCCTGA GGCCCATGAC CTACAAGGGC GCCGTGGACC
TGTCCTCACTT CCTGAAGGAG AAGGGCGGCC TGGAGGGCCT GATCCACTCC CAGAAGAGGC
AGGACATCCT GGACCTGTGG GTGTACCACA CCCAGGGCTA CTCCCCGAC TGGCAGAACT
ACACCCCCCG CCCCAGCATC AGGTTCCTCC TGACCTTCGG CTGGTGCTTC AAGCTGGTGC
CCGTGGAGCC CGAGAAGGTG GAGGAGGCCA ACGAGGGCGA GAACAACTGC CTGCTGCACC
```

CCATGTCCCA GCACGGCATC GAGGACCCCG AGAAGGAGGT GCTGGAGTGG AGGTTGACT
 CCAAGCTGGC CTTCCACCAC GTGGCCAGGG AGCTGCACCC CGAGTACTAC AAGGACTGCT
 AAAGCCCGGG C (SEQ ID NO:1).

- As can be discerned from comparing native to optimized codon usage in
 5 Figure 2A-B, the following codon usage for mammalian optimization is preferred:
 Met (ATG), Gly (GGC), Lys (AAG), Trp (TGG), Ser (TCC), Arg (AGG), Val (GTG),
 Pro (CCC), Thr (ACC), Glu (GAG); Leu (CTG), His (CAC), Ile (ATC), Asn (AAC),
 Cys (TGC), Ala (GCC), Gln (CAG), Phe (TTC) and Tyr (TAC). For an additional
 discussion relating to mammalian (human) codon optimization, see WO 97/31115
 10 (PCT/US97/02294), which is hereby incorporated by reference.

- The open reading frame for SEQ ID NO:1 above comprises an initiating
 methionine residue at nucleotides 12-14 and a "TAA" stop codon from nucleotides
 660-662. The open reading frame of SEQ ID NO:1 provides for a 216 amino acid
 HIV-1 Nef protein expressed through utilization of a codon optimized DNA vaccine
 15 vector. The 216 amino acid HIV-1 Nef (jfrl) protein is disclosed herein as SEQ ID
 NO:2, and as follows:

- Met Gly Gly Lys Trp Ser Lys Arg Ser Val Pro Gly Trp Ser Thr Val
 Arg Glu Arg Met Arg Arg Ala Glu Pro Ala Ala Asp Arg Val Arg Arg
 Thr Glu Pro Ala Ala Val Gly Val Gly Ala Val Ser Arg Asp Leu Glu
 20 Lys His Gly Ala Ile Thr Ser Ser Asn Thr Ala Ala Thr Asn Ala Asp
 Cys Ala Trp Leu Glu Ala Gln Glu Asp Glu Glu Val Gly Phe Pro Val
 Arg Pro Gln Val Pro Leu Arg Pro Met Thr Tyr Lys Gly Ala Val Asp
 Leu Ser His Phe Leu Lys Glu Lys Gly Gly Leu Glu Gly Leu Ile His
 Ser Gln Lys Arg Gln Asp Ile Leu Asp Leu Trp Val Tyr His Thr Gln
 25 Gly Tyr Phe Pro Asp Trp Gln Asn Tyr Thr Pro Gly Pro Gly Ile Arg
 Phe Pro Leu Thr Phe Gly Trp Cys Phe Lys Leu Val Pro Val Glu Pro
 Glu Lys Val Glu Glu Ala Asn Glu Gly Glu Asn Asn Cys Leu Leu His
 Pro Met Ser Gln His Gly Ile Glu Asp Pro Glu Lys Glu Val Leu Glu
 Trp Arg Phe Asp Ser Lys Leu Ala Phe His His Val Ala Arg Glu Leu
 30 His Pro Glu Tyr Tyr Lys Asp Cys (SEQ ID NO:2).

HIV-1 Nef is a 206 amino acid cytosolic protein which associates with the
 inner surface of the host cell plasma membrane through myristylation of Gly-2
 (Franchini et al., 1986, *Virology* 155: 593-599). While not all possible Nef functions
 have been elucidated, it has become clear that correct trafficking of Nef to the inner

plasma membrane promotes viral replication by altering the host intracellular environment to facilitate the early phase of the HIV-1 life cycle and by increasing the infectivity of progeny viral particles. In one aspect of the invention regarding codon-optimized, protein-modified polypeptides, either the DNA vaccine vector molecule or the HIV-1 nef construct is modified to contain a nucleotide sequence which encodes a heterologous leader peptide such that the amino terminal region of the expressed protein will contain the leader peptide. The diversity of function that typifies eukaryotic cells depends upon the structural differentiation of their membrane boundaries. To generate and maintain these structures, proteins must be transported from their site of synthesis in the endoplasmic reticulum to predetermined destinations throughout the cell. This requires that the trafficking proteins display sorting signals that are recognized by the molecular machinery responsible for route selection located at the access points to the main trafficking pathways. Sorting decisions for most proteins need to be made only once as they traverse their biosynthetic pathways since their final destination, the cellular location at which they perform their function, becomes their permanent residence. Maintenance of intracellular integrity depends in part on the selective sorting and accurate transport of proteins to their correct destinations. Defined sequence motifs exist in proteins which can act as 'address labels'. A number of sorting signals have been found associated with the cytoplasmic domains of membrane proteins. An effective induction of CTL responses often required sustained, high level endogenous expression of an antigen. In light of its diverse biological activities, vaccines composed of wild-type Nef could potentially have adverse effects on the host cells. As membrane-association via myristylation is an essential requirement for most of Nef's function, mutants lacking myristylation, by glycine-to-alanine change, change of the dileucine motif and/or by substitution with a tpa leader sequence as described herein, will be functionally defective, and therefore will have improved safety profile compared to wild-type Nef for use as an HIV-1 vaccine component.

In a preferred and exemplified embodiment of this portion of the invention, either the DNA vector or the HIV-1 nef nucleotide sequence is modified to include the human tissue-specific plasminogen activator (tPA) leader. As shown in Figure 1A-B for the DNA vector V1Jns, a DNA vector which may be utilized to practice the present invention may be modified by known recombinant DNA methodology to contain a leader signal peptide of interest, such that downstream

cloning of the modified HIV-1 protein of interest results in a nucleotide sequence which encodes a modified HIV-1 tPA/Nef protein. In the alternative, as noted above, insertion of a nucleotide sequence which encodes a leader peptide may be inserted into a DNA vector housing the open reading frame for the Nef protein of interest.

- 5 Regardless of the cloning strategy, the end result is a polynucleotide vaccine which comprises vector components for effective gene expression in conjunction with nucleotide sequences which encode a modified HIV-1 Nef protein of interest, including but not limited to a HIV-1 Nef protein which contains a leader peptide. The amino acid sequence of the human tPA leader utilized herein is as follows:
- 10 MDAMKRGGLCCVLLLCGAVFVSPSEISS (SEQ ID NO:19).

- It has been shown that myristylation of Gly-2 in conjunction with a dileucine motif in the carboxy region of the protein is essential for Nef-induced down regulation of CD4 (Aiken et al., 1994, *Cell* 76: 853-864) via endocytosis. It has also been shown that Nef expression promotes down regulation of MHC I (Schwartz et al., 15 1996, *Nature Medicine* 2(3): 338-342) via endocytosis. The present invention relates in part to DNA vaccines which encode modified Nef proteins altered in trafficking and/or functional properties. The modifications introduced into the DNA vaccines of the present invention include but are not limited to additions, deletions or substitutions to the nef open reading frame which results in the expression of a
- 20 modified Nef protein which includes an amino terminal leader peptide, modification or deletion of the amino terminal myristylation site, and modification or deletion of the dileucine motif within the Nef protein and which alter function within the infected host cell. Therefore, a central theme of the DNA molecules and DNA vaccines of the present invention is (1) host administration and intracellular delivery of a codon
- 25 optimized nef-based DNA vector vaccine; (2) expression of a modified Nef protein which is immunogenic in terms of eliciting both CTL and Th responses; and, (3) inhibiting or at least altering known early viral functions of Nef which have been shown to promote HIV-1 replication and load within an infected host.

- In another preferred and exemplified embodiment of the present invention, the
- 30 nef coding region is altered, resulting in a DNA vaccine which expresses a modified Nef protein wherein the amino terminal Gly-2 myristylation residue is either deleted or modified to express alternate amino acid residues.

In another preferred and exemplified embodiment of the present invention, the nef coding region is altered, resulting in a DNA vaccine which expresses a modified

Nef protein wherein the dileucine motif is either deleted or modified to express alternate amino acid residues.

Therefore, the present invention relates to an isolated DNA molecule, regardless of codon usage, which expresses a wild type or modified Nef protein as described herein, including but not limited to modified Nef proteins which comprise a deletion or substitution of Gly 2, a deletion or substitution of Leu 174 and Leu 175 and/or inclusion of a leader sequence.

The present invention also relates to a substantially purified protein expressed from the DNA polynucleotide vaccines of the present invention, especially the purified proteins set forth below as SEQ ID NOs: 2, 4, 6, and 8. These purified proteins may be useful as protein-based HIV vaccines.

In a specific embodiment of the invention as it relates DNA vaccines encoding modified forms of HIV-1, an open reading frame which encodes a Nef protein which comprises a tPA leader sequence fused to amino acid residue 6-216 of HIV-1 Nef (jfrl) is referred to herein as opt tpanef. The nucleotide sequence comprising the open reading frame of opt tpanef is disclosed herein as SEQ ID NO:3, as shown below:

CATGGATGCA ATGAAGAGAG GGCTCTGCTG TGTGCTGCTG CTGTGTGGAG CAGTCTTCGT
TTCGCCCAGC GAGATCTCCT CCAAGAGGTC CGTGCCCCGC TGGTCCACCG TGAGGGAGAG
GATGAGGAGG GCCGAGCCCG CCGCCGACAG GGTGAGGAGG ACCGAGCCCG CCGCCGTGGG
CGTGGGCGCC GTGTCCAGGG ACCTGGAGAA GCACGGCGCC ATCACCTCCT CCAACACCGC
CGCCACCAAC GCCGACTGCG CCTGGCTGGA GGCCAGGAG GACGAGGAGG TGGGCTTCCC
CGTGAGGCCC CAGGTGCCCC TGAGGCCCAT GACCTACAAG GGCGCCGTGG ACCTGTCCCA
CTTCCTGAAG GAGAAGGGCG GCCTGGAGGG CCTGATCCAC TCCCAGAAGA GGCAGGACAT
CCTGGACCTG TGGGTGTACC ACACCCAGGG CTACTTCCCC GACTGGCAGA ACTACACCCC
CGGCCCCGGC ATCAGGTTCC CCCTGACCTT CGGCTGGTGC TTCAAGCTGG TGCCCGTGGA
GCCCAGAGAAG GTGGAGGAGG CCAACGAGGG CGAGAACAAC TGCCTGCTGC ACCCATGTC
CCAGCACGGC ATCGAGGACC CCGAGAAGGA GGTGCTGGAG TGGAGGTTTC ACTCCAAGCT
GGCCTTCCAC CACGTGGCCA GGGAGCTGCA CCCCAGGTAC TACAAGGACT GCTAAAGCC
(SEQ ID NO:3).

The open reading frame for SEQ ID NO:3 comprises an initiating methionine

residue at nucleotides 2-4 and a "TAA" stop codon from nucleotides 713-715. The open reading frame of SEQ ID NO:3 provides for a 237 amino acid HIV-1 Nef protein which comprises a tPA leader sequence fused to amino acids 6-216 of HIV-1 Nef, including the dileucine motif at amino acid residues 174 and 175. This 237 amino acid tPA/Nef (jfrl) fusion protein is disclosed herein as SEQ ID NO:4, and is shown as follows:

Met Asp Ala Met Lys Arg Gly Leu Cys Cys Val Leu Leu Leu Cys Gly
 Ala Val Phe Val Ser Pro Ser Glu Ile Ser Ser Lys Arg Ser Val Pro
 Gly Trp Ser Thr Val Arg Glu Arg Met Arg Arg Ala Glu Pro Ala Ala
 10 Asp Arg Val Arg Arg Thr Glu Pro Ala Ala Val Gly Val Gly Ala Val
 Ser Arg Asp Leu Glu Lys His Gly Ala Ile Thr Ser Ser Asn Thr Ala
 Ala Thr Asn Ala Asp Cys Ala Trp Leu Glu Ala Gln Glu Asp Glu Glu
 Val Gly Phe Pro Val Arg Pro Gln Val Pro Leu Arg Pro Met Thr Tyr
 Lys Gly Ala Val Asp Leu Ser His Phe Leu Lys Glu Lys Gly Gly Leu
 15 Glu Gly Leu Ile His Ser Gln Lys Arg Gln Asp Ile Leu Asp Leu Trp
 Val Tyr His Thr Gln Gly Tyr Phe Pro Asp Trp Gln Asn Tyr Thr Pro
 Gly Pro Gly Ile Arg Phe Pro Leu Thr Phe Gly Trp Cys Phe Lys Leu
 Val Pro Val Glu Pro Glu Lys Val Glu Glu Ala Asn Glu Gly Glu Asn
 Asn Cys Leu Leu His Pro Met Ser Gln His Gly Ile Glu Asp Pro Glu
 20 Lys Glu Val Leu Glu Trp Arg Phe Asp Ser Lys Leu Ala Phe His His
 Val Ala Arg Glu Leu His Pro Glu Tyr Tyr Lys Asp Cys (SEQ ID NO:4).

Therefore, this exemplified Nef protein, Opt tPA-Nef, contains both a tPA leader sequence as well as deleting the myristylation site of Gly-2A DNA molecule encoding HIV-1 Nef from the HIV-1 jfrl isolate wherein the codons are optimized for expression in a mammalian system such as a human.

In another specific embodiment of the present invention, a DNA molecule is disclosed which encodes optimized HIV-1 Nef wherein the open reading frame codes for modifications at the amino terminal myristylation site (Gly-2 to Ala-2) and substitution of the Leu-174-Leu-175 dileucine motif to Ala-174-Ala-175. This open reading frame is herein described as opt nef (G2A,LLAA) and is disclosed as SEQ ID NO:5, which comprises an initiating methionine residue at nucleotides 12-14 and a "TAA" stop codon from nucleotides 660-662. The nucleotide sequence of this codon optimized version of HIV-1 jfrl nef gene with the above mentioned modifications is disclosed herein as SEQ ID NO:5, as follows:

GATCTGCCAC CATGGCCGGC AAGTGGTCCA AGAGGTCCGT GCCCGGCTGG TCCACCGTGA
 GGGAGAGGAT GAGGAGGGCC GAGCCCGCCG CCGACAGGGT GAGGAGGACC GAGCCCGCCC
 CCGTGGGCGT GGGCGCCGTG TCCAGGGACC TGGAGAAGCA CGGCGCCATC ACCTCCTCCA
 ACACCGCCGC CACCAACGCC GACTGCGCCT GGCTGGAGGC CCAGGAGGAC GAGGAGGTGG
 5 GCTTCCCCGT GAGGCCCCAG GTGCCCCTGA GGCCCATGAC CTACAAGGGC GCCGTGGACC
 TGTCCCACTT CCTGAAGGAG AAGGGCGGCC TGGAGGGCCT GATCCACTCC CAGAAGAGGC
 AGGACATCCT GGACCTGTGG GTGTACCACA CCCAGGGCTA CTTCCCCGAC TGGCAGAACT
 ACACCCCCGG CCCCGGCATC AGGTTCCTCC TGACCTTCGG CTGGTGCTTC AAGCTGGTGC
 CCGTGGAGCC CGAGAAGGTG GAGGAGGCCA ACGAGGGCGA GAACAACCTGC GCCGCCACC
 10 CCATGTCCCA GCACGGCATC GAGGACCCCG AGAAGGAGGT GCTGGAGTGG AGGTTCGACT
 CCAAGCTGGC CTTCCACCAC GTGGCCAGGG AGCTGCACCC CGAGTACTAC AAGGACTGCT
 AAAGCCCGGG C (SEQ ID NO:5).

The open reading frame of SEQ ID NO:5 encodes Nef (G2A,LLAA), disclosed herein as SEQ ID NO:6, as follows:

15 Met Ala Gly Lys Trp Ser Lys Arg Ser Val Pro Gly Trp Ser Thr Val
 Arg Glu Arg Met Arg Arg Ala Glu Pro Ala Ala Asp Arg Val Arg Arg
 Thr Glu Pro Ala Ala Val Gly Val Gly Ala Val Ser Arg Asp Leu Glu
 Lys His Gly Ala Ile Thr Ser Ser Asn Thr Ala Ala Thr Asn Ala Asp
 Cys Ala Trp Leu Glu Ala Gln Glu Asp Glu Glu Val Gly Phe Pro Val
 20 Arg Pro Gln Val Pro Leu Arg Pro Met Thr Tyr Lys Gly Ala Val Asp
 Leu Ser His Phe Leu Lys Glu Lys Gly Gly Leu Glu Gly Leu Ile His
 Ser Gln Lys Arg Gln Asp Ile Leu Asp Leu Trp Val Tyr His Thr Gln
 Gly Tyr Phe Pro Asp Trp Gln Asn Tyr Thr Pro Gly Pro Gly Ile Arg
 Phe Pro Leu Thr Phe Gly Trp Cys Phe Lys Leu Val Pro Val Glu Pro
 25 Glu Lys Val Glu Glu Ala Asn Glu Gly Glu Asn Asn Cys Ala Ala His
 Pro Met Ser Gln His Gly Ile Glu Asp Pro Glu Lys Glu Val Leu Glu
 Trp Arg Phe Asp Ser Lys Leu Ala Phe His His Val Ala Arg Glu Leu
 His Pro Glu Tyr Tyr Lys Asp Cys Ser (SEQ ID NO:6).

30 An additional embodiment of the present invention relates to another DNA
 molecule encoding optimized HIV-1 Nef wherein the amino terminal myristylation
 site and dileucine motif have been deleted, as well as comprising a tPA leader peptide.
 This DNA molecule, opt tpanef (LLAA) comprises an open reading frame which
 encodes a Nef protein containing a tPA leader sequence fused to amino acid residue
 6-216 of HIV-1 Nef (jfrl), wherein Leu-174 and Leu-175 are substituted with Ala-174

and Ala-175 (Ala-195 and Ala-196 in this tPA-based fusion protein). The nucleotide sequence comprising the open reading frame of opt tpanef (LLAA) is disclosed herein as SEQ ID NO:7, as shown below:

```

CATGGATGCA ATGAAGAGAG GGCTCTGCTG TGTGCTGCTG CTGTGTGGAG CAGTCTTCGT
5  TTCGCCCAGC GAGATCTCCT CCAAGAGGTC CGTGCCCGGC TGGTCCACCG TGAGGGAGAG
  GATGAGGAGG GCCGAGCCCG CCGCCGACAG GGTGAGGAGG ACCGAGCCCG CCGCCGTGGG
  CGTGGGCGCC GTGTCCAGGG ACCTGGAGAA GCACGGCGCC ATCACCTCCT CCAACACCGC
  CGCCACCAAC GCCGACTGCG CCTGGCTGGA GGCCAGGAG GACGAGGAGG TGGGCTTCCC
  CGTGAGGCCC CAGGTGCCCC TGAGGCCCAT GACCTACAAG GCGCCGTGG ACCTGTCCCA
10  CTTCTGAAG GAGAAGGGCG GCCTGGAGGG CCTGATCCAC TCCCAGAAGA GGCAGGACAT
  CCTGGACCTG TGGGTGTACC ACACCCAGGG CTACTTCCCC GACTGGCAGA ACTACACCCC
  CGGCCCCGGC ATCAGGTTCC CCCTGACCTT CGGCTGGTGC TTCAAGCTGG TGCCCGTGGA
  GCCCGAGAAG GTGGAGGAGG CCAACGAGGG CGAGAACAAC TCGCCGCCC ACCCATGTC
  CCAGCACGGC ATCGAGGACC CCGAGAAGGA GGTGCTGGAG TGGAGGTTCC ACTCCAAGCT
15  GGCCTTCCAC CACGTGGCCA GGGAGCTGCA CCCCAGTAC TACAAGGACT GCTAAAGCCC
  (SEQ ID NO:7).

```

The open reading frame of SEQ ID NO:7 encoding tPA-Nef (LLAA), disclosed herein as SEQ ID NO:8, is as follows:

```

Met Asp Ala Met Lys Arg Gly Leu Cys Cys Val Leu Leu Leu Cys Gly
20  Ala Val Phe Val Ser Pro Ser Glu Ile Ser Ser Lys Arg Ser Val Pro
  Gly Trp Ser Thr Val Arg Glu Arg Met Arg Arg Ala Glu Pro Ala Ala
  Asp Arg Val Arg Arg Thr Glu Pro Ala Ala Val Gly Val Gly Ala Val
  Ser Arg Asp Leu Glu Lys His Gly Ala Ile Thr Ser Ser Asn Thr Ala
  Ala Thr Asn-Ala Asp Cys Ala Trp Leu Glu Ala Gln Glu Asp Glu Glu
25  Val Gly Phe Pro Val Arg Pro Gln Val Pro Leu Arg Pro Met Thr Tyr
  Lys Gly Ala Val Asp Leu Ser His Phe Leu Lys Glu Lys Gly Gly Leu
  Glu Gly Leu Ile His Ser Gln Lys Arg Gln Asp Ile Leu Asp Leu Trp
  Val Tyr His Thr Gln Gly Tyr Phe Pro Asp Trp Gln Asn Tyr Thr Pro
  Gly Pro Gly Ile Arg Phe Pro Leu Thr Phe Gly Trp Cys Phe Lys Leu
30  Val Pro Val Glu Pro Glu Lys Val Glu Glu Ala Asn Glu Gly Glu Asn
  Asn Cys Ala Ala His Pro Met Ser Gln His Gly Ile Glu Asp Pro Glu
  Lys Glu Val Leu Glu Trp Arg Phe Asp Ser Lys Leu Ala Phe His His
  Val Ala Arg Glu Leu His Pro Glu Tyr Tyr Lys Asp Cys (SEQ ID NO:8).

```

The present invention also relates in part to any DNA molecule, regardless of

codon usage, which expresses a wild type or modified Nef protein as described herein, including but not limited to modified Nef proteins which comprise a deletion or substitution of Gly 2, a deletion of substitution of Leu 174 and Leu 175 and/or inclusion of a leader sequence. Therefore, partial or fully codon optimized DNA vaccine expression vector constructs are preferred since such constructs should result in increased host expression. However, it is within the scope of the present invention to utilize "non-codon optimized" versions of the constructs disclosed herein, especially modified versions of HIV Nef which are shown to promote a substantial cellular immune response subsequent to host administration.

10 The DNA backbone of the DNA vaccines of the present invention are preferably DNA plasmid expression vectors. DNA plasmid expression vectors are well known in the art and the present DNA vector vaccines may be comprised of any such expression backbone which contains at least a promoter for RNA polymerase transcription, and a transcriptional terminator 3' to the HIV nef coding sequence. In 15 one preferred embodiment, the promoter is the Rous sarcoma virus (RSV) long terminal repeat (LTR) which is a strong transcriptional promoter. A more preferred promoter is the cytomegalovirus promoter with the intron A sequence (CMV-intA). A preferred transcriptional terminator is the bovine growth hormone terminator. In addition, to assist in large scale preparation of an HIV nef DNA vector vaccine, an 20 antibiotic resistance marker is also preferably included in the expression vector. Ampicillin resistance genes, neomycin resistance genes or any other pharmaceutically acceptable antibiotic resistance marker may be used. In a preferred embodiment of this invention, the antibiotic resistance gene encodes a gene product for neomycin resistance. Further, to aid in the high level production of the pharmaceutical by 25 fermentation in prokaryotic organisms, it is advantageous for the vector to contain an origin of replication and be of high copy number. Any of a number of commercially available prokaryotic cloning vectors provide these benefits. In a preferred embodiment of this invention, these functionalities are provided by the commercially available vectors known as pUC. It is desirable to remove non-essential DNA 30 sequences. Thus, the lacZ and lacI coding sequences of pUC are removed in one embodiment of the invention.

DNA expression vectors exemplified herein are also disclosed in PCT International Application No. PCT/US94/02751, International Publication No. WO 94/21797, hereby incorporated by reference. A first DNA expression vector

is the expression vector pnRSV, wherein the rous sarcoma virus (RSV) long terminal repeat (LTR) is used as the promoter. A second embodiment relates to plasmid V1, a mutated pBR322 vector into which the CMV promoter and the BGH transcriptional terminator is cloned. Another embodiment regarding DNA vector backbones relates to plasmid V1J. Plasmid V1J is derived from plasmid V1 and removes promoter and transcription termination elements in order to place them within a more defined context, create a more compact vector, and to improve plasmid purification yields. Therefore, V1J also contains the CMVintA promoter and (BGH) transcription termination elements which control the expression of the HIV nef-based genes disclosed herein. The backbone of V1J is provided by pUC18. It is known to produce high yields of plasmid, is well-characterized by sequence and function, and is of minimum size. The entire *lac* operon was removed and the remaining plasmid was purified from an agarose electrophoresis gel, blunt-ended with the T4 DNA polymerase, treated with calf intestinal alkaline phosphatase, and ligated to the CMVintA/BGH element. In another DNA expression vector, the ampicillin resistance gene is removed from V1J and replaced with a neomycin resistance gene, to generate V1Jneo. A DNA expression vector specifically exemplified herein is V1Jns, which is the same as V1J except that a unique SfiI restriction site has been engineered into the single KpnI site at position 2114 of V1J-neo. The incidence of SfiI sites in human genomic DNA is very low (approximately 1 site per 100,000 bases). Thus, this vector allows careful monitoring for expression vector integration into host DNA, simply by SfiI digestion of extracted genomic DNA. Another DNA expression vector for use as the backbone to the HIV-1 nef-based DNA vaccines of the present invention is V1R. In this vector, as much non-essential DNA as possible is "trimmed" from the vector to produce a highly compact vector. This vector is a derivative of V1Jns. This vector allows larger inserts to be used, with less concern that undesirable sequences are encoded and optimizes uptake by cells when the construct encoding specific influenza virus genes is introduced into surrounding tissue.

It will be evident upon review of the teaching within this specification that numerous vector/Nef antigen constructs may be generated. While the exemplified constructs (V1Jns/nef, V1Jns/tpanef, V1Jns/tpanef(LLAA) and V1Jns/(G2A,LLAA) are preferred, any number of vector/Nef antigen combinations are within the scope of the present invention, especially wild type or modified Nef proteins which comprise a deletion or substitution of Gly 2, a deletion or substitution of Leu 174 and Leu 175

and/or inclusion of a leader sequence. Therefore, the present invention especially relates to DNA vaccines and a pharmaceutically active vaccine composition which contains this DNA vector vaccine, and the use as prophylactic and/or therapeutic vaccine for host immunization, preferably human host immunization, against an HIV infection or to combat an existing HIV condition. These DNA vaccines are represented by codon optimized DNA molecules encoding HIV-1 Nef of biologically active Nef modifications or Nef-containing fusion proteins which are ligated within an appropriate DNA plasmid vector, with or without a nucleotide sequence encoding a functional leader peptide. DNA vaccines of the present invention include but in no way are limited to codon optimized DNA molecules encoding HIV-1 Nef of biologically active Nef modifications or Nef-containing fusion proteins ligated in DNA vectors V1, V1J (SEQ ID NO:14), V1Jneo (SEQ ID NO:15), V1Jns (Figure 1A, SEQ ID NO:16), V1R (SEQ ID NO:26), or any of the aforementioned vectors wherein a nucleotide sequence encoding a leader peptide, preferably the human tPA leader, is fused directly downstream of the CMV-intA promoter, including but not limited to V1Jns-tpa, as shown in Figure 1B and SEQ ID NO:19. Especially preferred DNA vaccines of the present invention include as V1Jns/nef, V1Jns/tpanef, V1Jns/tpanef(LLAA) and V1Jns/(G2A,LLAA), as exemplified in Example Section 2.

The DNA vector vaccines of the present invention may be formulated in any pharmaceutically effective formulation for host administration. Any such formulation may be, for example, a saline solution such as phosphate buffered saline (PBS). It will be useful to utilize pharmaceutically acceptable formulations which also provide long-term stability of the DNA vector vaccines of the present invention. During storage as a pharmaceutical entity, DNA plasmid vaccines undergo a physiochemical change in which the supercoiled plasmid converts to the open circular and linear form. A variety of storage conditions (low pH, high temperature, low ionic strength) can accelerate this process. Therefore, the removal and/or chelation of trace metal ions (with succinic or malic acid, or with chelators containing multiple phosphate ligands) from the DNA plasmid solution, from the formulation buffers or from the vials and closures, stabilizes the DNA plasmid from this degradation pathway during storage. In addition, inclusion of non-reducing free radical scavengers, such as ethanol or glycerol, are useful to prevent damage of the DNA plasmid from free radical production that may still occur, even in apparently demetalated solutions. Furthermore, the buffer type, pH, salt concentration, light exposure, as well as the

type of sterilization process used to prepare the vials, may be controlled in the formulation to optimize the stability of the DNA vaccine. Therefore, formulations that will provide the highest stability of the DNA vaccine will be one that includes a demetalated solution containing a buffer (phosphate or bicarbonate) with a pH in the range of 7-8, a salt (NaCl, KCl or LiCl) in the range of 100-200 mM, a metal ion chelator (e.g., EDTA, diethylenetriaminepenta-acetic acid (DTPA), malate, inositol hexaphosphate, tripolyphosphate or polyphosphoric acid), a non-reducing free radical scavenger (e.g. ethanol, glycerol, methionine or dimethyl sulfoxide) and the highest appropriate DNA concentration in a sterile glass vial, packaged to protect the highly purified, nuclease free DNA from light. A particularly preferred formulation which will enhance long term stability of the DNA vector vaccines of the present invention would comprise a Tris-HCl buffer at a pH from about 8.0 to about 9.0; ethanol or glycerol at about 3% w/v; EDTA or DTPA in a concentration range up to about 5 mM; and NaCl at a concentration from about 50 mM to about 500 mM. The use of such stabilized DNA vector vaccines and various alternatives to this preferred formulation range is described in detail in PCT International Application No. PCT/US97/06655, PCT International Publication No. WO 97/40839, which is hereby incorporated by reference.

The DNA vector vaccines of the present invention may, in addition to generating a strong CTL-based immune response, provide for a measurable humoral response subsequent immunization. This response may occur with or without the addition of adjuvant to the respective vaccine formulation. To this end, the DNA vector vaccines of the present invention may also be formulated with an adjuvant or adjuvants which may increase immunogenicity of the DNA polynucleotide vaccines of the present invention. A number of these adjuvants are known in the art and are available for use in a DNA vaccine, including but not limited to particle bombardment using DNA-coated gold beads, co-administration of DNA vaccines with plasmid DNA expressing cytokines, chemokines, or costimulatory molecules, formulation of DNA with cationic lipids or with experimental adjuvants such as saponin, monophosphoryl lipid A or other compounds which increase immunogenicity of the DNA vaccine. One preferred adjuvant for use in the DNA vector vaccines of the present invention are one or more forms of an aluminum phosphate-based adjuvant. Aluminum phosphate is known in the art for use with live, killed or subunit vaccines, but is only recently

disclosed as a useful adjuvant in DNA vaccine formulations. The artisan may alter the ratio of DNA to aluminum phosphate to provide for an optimal immune response. In addition, the aluminum phosphate-based adjuvant possesses a molar PO_4/Al ratio of approximately 0.9, and may again be altered by the skilled artisan to provide for an optimal immune response. An additional mineral-based adjuvant may be generated from one or more forms of a calcium phosphate. These mineral-based adjuvants are useful in increasing humoral responses to DNA vaccination without imparting a negative effect on an appropriate cellular immune response. Complete guidance for use of these mineral-based compounds for use as DNA vaccines adjuvants are disclosed in PCT International Application No. PCT/US98/02414, PCT International Publication No. WO 98/35562, which are hereby incorporated by reference in their entirety. Another preferred adjuvant is a non-ionic block copolymer which shows adjuvant activity with DNA vaccines. The basic structure comprises blocks of polyoxyethylene (POE) and polyoxypropylene (POP) such as a POE-POP-POE block copolymer. Newman et al. (1998, *Critical Reviews in Therapeutic Drug Carrier Systems* 15(2): 89-142) review a class of non-ionic block copolymers which show adjuvant activity. The basic structure comprises blocks of polyoxyethylene (POE) and polyoxypropylene (POP) such as a POE-POP-POE block copolymer. Newman et al. *id.*, disclose that certain POE-POP-POE block copolymers may be useful as adjuvants to an influenza protein-based vaccine, namely higher molecular weight POE-POP-POE block copolymers containing a central POP block having a molecular weight of over about 9000 daltons to about 20,000 daltons and flanking POE blocks which comprise up to about 20% of the total molecular weight of the copolymer (see also U.S. Reissue Patent No. 36,665, U.S. Patent No. 5,567,859, U.S. Patent No. 5,691,387, U.S. Patent No. 5,696,298 and U.S. Patent No. 5,990,241, all issued to Emanuele, et al., regarding these POE-POP-POE block copolymers). WO 96/04932 further discloses higher molecular weight POE/POP block copolymers which have surfactant characteristics and show biological efficacy as vaccine adjuvants. The above cited references within this paragraph are hereby incorporated by reference in their entirety. It is therefore within the purview of the skilled artisan to utilize available adjuvants which may increase the immune response of the polynucleotide vaccines of the present invention in comparison to administration of a non-adjuvanted polynucleotide vaccine.

The DNA vector vaccines of the present invention are administered to the host by any means known in the art, such as enteral and parenteral routes. These routes of delivery include but are not limited to intramuscular injection, intraperitoneal injection, intravenous injection, inhalation or intranasal delivery, oral delivery, sublingual administration, subcutaneous administration, transdermal administration, transcutaneous administration, percutaneous administration or any form of particle bombardment, such as a biolistic device such as a "gene gun" or by any available needle-free injection device. The preferred methods of delivery of the HIV-1 Nef-based DNA vaccines disclosed herein are intramuscular injection and needle-free injection. An especially preferred method is intramuscular delivery.

The amount of expressible DNA to be introduced to a vaccine recipient will depend on the strength of the transcriptional and translational promoters used in the DNA construct, and on the immunogenicity of the expressed gene product. In general, an immunologically or prophylactically effective dose of about 1 μ g to greater than about 20 mg, and preferably in doses from about 1 mg to about 5 mg is administered directly into muscle tissue. As noted above, subcutaneous injection, intradermal introduction, impression through the skin, and other modes of administration such as intraperitoneal, intravenous, inhalation and oral delivery are also contemplated. It is also contemplated that booster vaccinations are to be provided in a fashion which optimizes the overall immune response to the Nef-based DNA vector vaccines of the present invention.

The aforementioned polynucleotides, when directly introduced into a vertebrate *in vivo*, express the respective HIV-1 Nef protein within the animal and in turn induce a cytotoxic T lymphocyte (CTL) response within the host to the expressed Nef antigen. To this end, the present invention also relates to methods of using the HIV-1 Nef-based polynucleotide vaccines of the present invention to provide effective immunoprophylaxis, to prevent establishment of an HIV-1 infection following exposure to this virus, or as a post-HIV infection therapeutic vaccine to mitigate the acute HIV-1 infection so as to result in the establishment of a lower virus load with beneficial long term consequences. As noted above, the present invention contemplates a method of administration or use of the DNA nef-based vaccines of the present invention using any of the known routes of introducing polynucleotides into living tissue to induce expression of proteins.

Therefore, the present invention provides for methods of using a DNA nef-

based vaccine utilizing the various parameters disclosed herein as well as any additional parameters known in the art, which, upon introduction into mammalian tissue induces *in vivo*, intracellular expression of these DNA nef-based vaccines. This intracellular expression of the Nef-based immunogen induces a CTL and humoral
 5 response which provides a substantial level of protection against an existing HIV-1 infection or provides a substantial level of protection against a future infection in a presently uninfected host.

The following examples are provided to illustrate the present invention without, however, limiting the same hereto.

10

EXAMPLE 1

Vaccine Vectors

V1 – Vaccine vector V1 was constructed from pCMVIE-AKI-DHFR (Whang et al., 1987, *J. Virol.* 61: 1796). The AKI and DHFR genes were removed by cutting
 15 the vector with EcoRI and self-ligating. This vector does not contain intron A in the CMV promoter, so it was added as a PCR fragment that had a deleted internal SacI site [at 1855 as numbered in Chapman, et al., (1991, *Nuc. Acids Res.* 19: 3979)]. The template used for the PCR reactions was pCMVintA-Lux, made by ligating the HindIII and NheI fragment from pCMV6a120 (see Chapman et al., *ibid.*), which
 20 includes hCMV-IE1 enhancer/promoter and intron A, into the HindIII and XbaI sites of pBL3 to generate pCMVIntBL. The 1881 base pair luciferase gene fragment (HindIII-SmaI Klenow filled-in) from RSV-Lux (de Wet et al., 1987, *Mol. Cell Biol.* 7: 725) was ligated into the Sall site of pCMVIntBL, which was Klenow filled-in and phosphatase treated. The primers that spanned intron A are: 5' primer:
 25 5'-CTATATAAGCAGAGCTCGTTTAG-3' (SEQ ID NO:10); 3' primer:
 5'-GTAGCAAAGATCTAAGGACGGTGACTGCAG-3' (SEQ ID NO:11). The primers used to remove the SacI site are: sense primer, 5'-GTATGTGTCTG AAAATGAGC GTGGAGATTGGGCTCGCAC-3' (SEQ ID NO:12) and the antisense primer, 5'-GTGCGAGCCCAATCTCCACGCTCATTTTCAGAC
 30 ACATAC-3' (SEQ ID NO:13). The PCR fragment was cut with Sac I and Bgl II and inserted into the vector which had been cut with the same enzymes.

VIIJ – Vaccine vector VIIJ was generated to remove the promoter and transcription termination elements from vector V1 in order to place them within a more defined context, create a more compact vector, and to improve plasmid

purification yields. VIJ is derived from vectors V1 and pUC18, a commercially available plasmid. V1 was digested with SspI and EcoRI restriction enzymes producing two fragments of DNA. The smaller of these fragments, containing the CMVintA promoter and Bovine Growth Hormone (BGH) transcription termination elements which control the expression of heterologous genes, was purified from an agarose electrophoresis gel. The ends of this DNA fragment were then "blunted" using the T4 DNA polymerase enzyme in order to facilitate its ligation to another "blunt-ended" DNA fragment. pUC18 was chosen to provide the "backbone" of the expression vector. It is known to produce high yields of plasmid, is well-characterized by sequence and function, and is of small size. The entire *lac* operon was removed from this vector by partial digestion with the HaeII restriction enzyme. The remaining plasmid was purified from an agarose electrophoresis gel, blunt-ended with the T4 DNA polymerase treated with calf intestinal alkaline phosphatase, and ligated to the CMVintA/BGH element described above. Plasmids exhibiting either of two possible orientations of the promoter elements within the pUC backbone were obtained. One of these plasmids gave much higher yields of DNA in *E. coli* and was designated VIJ. This vector's structure was verified by sequence analysis of the junction regions and was subsequently demonstrated to give comparable or higher expression of heterologous genes compared with V1. The nucleotide sequence of VIJ is as follows:

```

TCGCGCGTTT CCGTGATGAC GGTGAAAACC TCTGACACAT GCAGTCCCG GAGACGGTCA
CAGCTTGTCT GTAAGCGGAT GCCGGGAGCA GACAAGCCCG TCAGGGCGCG TCAGCGGGTG
TTGGCGGGTG TCGGGGCTGG CTTAACTATG CGGCATCAGA GCAGATTGTA CTGAGAGTGC
ACCATATGCG GTGTGAAATA CCGCACAGAT GCGTAAGGAG AAAATACCGC ATCAGATTGG
25 CTATTGGCCA TTGCATACGT TGTATCCATA TCATAATATG TACATTTATA TTGGCTCATG
TCCAACATTA CCGCCATGTT GACATTGATT ATTGACTAGT TATTAATAGT AATCAATTAC
GGGGTCATTA GTTCATAGCC CATATATGGA GTTCCGCGTT ACATAACTTA CGGTAAATGG
CCCGCCTGGC TGACCGCCCA ACGACCCCG CCCATTGACG TCAATAATGA CGTATGTTCC
CATAGTAACG CCAATAGGGA CTTTCCATTG ACGTCAATGG GTGGAGTATT TACGGTAAAC
30 TGCCCACTTG GCAGTACATC AAGTGTATCA TATGCCAAGT ACGCCCCCTA TTGACGTCAA
TGACGGTAAA TGGCCCGCCT GGCATTATGC CCAGTACATG ACCTTATGGG ACTTTCCTAC
TTGGCAGTAC ATCTACGTAT TAGTCATCGC TATTACCATG GTGATGCGGT TTTGGCAGTA
CATCAATGGG CGTGGATAGC GGTTTGACTC ACGGGGATTT CCAAGTCTCC ACCCCATTGA
CGTCAATGGG AGTTTGTTTT GGCACCAAAA TCAACGGGAC TTTCCAAAAT GTCGTAACAA

```

CTCCGCCCCA TTGACGCAAA TGGGCGGTAG GCGTGTACGG TGGGAGGTCT ATATAAGCAG
 AGCTCGTTTA GTGAACCGTC AGATCGCCTG GAGACGCCAT CCACGCTGTT TTGACCTCCA
 TAGAAGACAC CGGGACCGAT CCAGCCTCCG CGGCCGGGAA CGGTGCATTG GAACGCGGAT
 TCCCCGTGCC AAGAGTGACG TAAGTACCGC CTATAGAGTC TATAGGCCCA CCCCTTGGC
 5 TTCTTATGCA TGCTATACTG TTTTGGCTT GGGGTCTATA CACCCCGCT TCCTCATGTT
 ATAGGTGATG GTATAGCTTA GCCTATAGGT GTGGGTATT GACCATTATT GACCACTCCC
 CTATTGGTGA CGATACTTTC CATTACTAAT CCATAACATG GCTCTTTGCC ACAACTCTCT
 TTATTGGCTA TATGCCAATA CACTGTCTT CAGAGACTGA CACGGACTCT GTATTTTAC
 AGGATGGGGT CTCATTTATT ATTTACAAAT TCACATATAC AACACCACCG TCCCCAGTGC
 10 CCGCAGTTTT TATTAAACAT AACGTGGGAT CTCCACGCGA ATCTCGGGTA CGTGTTCGGG
 ACATGGGCTC TTCTCCGGTA GCGGCGGAGC TTCTACATCC GAGCCCTGCT CCCATGCCTC
 CAGCGACTCA TGGTCGCTCG GCAGCTCCTT GTCCTAACA GTGGAGGCCA GACTTAGGCA
 CAGCACGATG CCCACCACCA CCAGTGTGCC GCACAAGGCC GTGGCGGTAG GGTATGTGTC
 TGAAAATGAG CTCGGGGAGC GGGCTTGAC CGCTGACGCA TTTGGAAGAC TTAAGGCAGC
 15 GGCAGAAGAA GATGCAGGCA GCTGAGTTGT TGTGTTCTGA TAAGAGTCAG AGGTAACCTC
 CGTTGCGGTG CTGTTAACGG TGGAGGGCAG TGTAGTCTGA GCAGTACTCG TTGCTGCCGC
 GCGCGCCACC AGACATAATA GCTGACAGAC TAACAGACTG TTCCTTTCCA TGGGTCTTTT
 CTGCAGTCAC CGTCCCTAGA TCTGCTGTGC CTTCTAGTTG CCAGCCATCT GTTGTTCGCC
 CCTCCCCCGT GCCTTCCTTG ACCCTGGAAG GTGCCACTCC CACTGTCTT TCCTAATAAA
 20 ATGAGGAAAT TGCATCGCAT TGTCTGAGTA GGTGTCATTC TATTCTGGGG GGTGGGGTGG
 GGCAGCACAG CAAGGGGGAG GATTGGGAAG ACAATAGCAG GCATGCTGGG GATGCGGTGG
 GCTCTATGGG TACCCAGGTG CTGAAGAATT GACCCGGTTC CTCCTGGGCC AGAAAGAAGC
 AGGCACATCC CCTTCTCTGT GACACACCTT GTCCACGCCC CTGGTTCTTA GTTCCAGCCC
 CACTCATAGG ACATCATAG CTCAGGAGGG CTCCGCCTTC AATCCCACCC GCTAAAGTAC
 25 TTGGAGCGGT CTCTCCCTCC CTCATCAGCC CACCAAACCA AACCTAGCCT CCAAGAGTGG
 GAAGAAATTA AAGCAAGATA GGCTATTAAG TGCAGAGGGA GAGAAAATGC CTCCAACATG
 TGAGGAAGTA ATGAGAGAAA TCATAGAATT TCTTCCGCTT CCTCGCTCAC TGACTCGCTG
 CGCTCGGTG TTCGGCTGCG GCGAGCGGTA TCAGCTCACT CAAAGGCGGT AATACGGTTA
 TCCACAGAAT CAGGGGATAA CGCAGGAAAG AACATGTGAG CAAAAGGCCA GCAAAGGCC
 30 AGGAACCGTA AAAAGGCCGC GTTGTGGCG TTTTTCATA GGCTCCGCCC CCCTGACGAG
 CATCACAAAA ATCAGCGCTC AAGTCAGAGG TGGCGAAACC CGACAGGACT ATAAAGATAC
 CAGGCGTTTC CCCCTGGAAG CTCCCTCGTG CGCTCTCTTG TTCCGACCCT GCCGCTTACC
 GGATACCTGT CCGCCTTTCT CCCTTCGGGA AGCGTGGCGC TTTCTCAATG CTCACGCTGT
 AGGTATCTCA GTTCGGTGTA GGTCGTTTCG TCCAAGCTGG GCTGTGTGCA CGAACCCCCC

GTTCAGCCCG ACCGCTGCGC CTTATCCGGT AACTATCGTC TTGAGTCAA CCCGGTAAGA
 CACGACTTAT CGCCACTGGC AGCAGCCACT GGTAACAGGA TTAGCAGAGC GAGGTATGTA
 GCGGGTGCTA CAGAGTTCTT GAAGTGGTGG CCTAACTACG GCTACACTAG AAGGACAGTA
 TTTGGTATCT GCGCTCTGCT GAAGCCAGTT ACCTTCGGAA AAAGAGTTGG TAGCTCTTGA
 5 TCCGGCAAAC AAACCACCGC TGGTAGCGGT GGTTTTTTTG TTTGCAAGCA GCAGATTACG
 CGCAGAAAAA AAGGATCTCA AGAAGATCCT TTGATCTTTT CTACGGGGTC TGACGCTCAG
 TGGAACGAAA ACTCACGTTA AGGGATTTTG GTCATGAGAT TATCAAAAAG GATCTTCACC
 TAGATCCTTT TAAATTAAAA ATGAAGTTTT AAATCAATCT AAAGTATATA TGAGTAAACT
 TGGTCTGACA GTTACCAATG CTTAATCAGT GAGGCACCTA TCTCAGCGAT CTGTCTATTT
 10 CGTTCATCCA TAGTTGCCTG ACTCCCCGTC GTGTAGATAA CTACGATACG GGAGGGCTTA
 CCATCTGGCC CCAGTGCTGC AATGATACCG CGAGACCCAC GCTCACCAGC TCCAGATTTA
 TCAGCAATAA ACCAGCCAGC CGGAAGGGCC GAGCGCAGAA GTGGTCCTGC AACTTTATCC
 GCCTCCATCC AGTCTATTAA TTGTTGCCGG GAAGCTAGAG TAAGTAGTTC GCCAGTTAAT
 AGTTTGCGCA ACGTTGTTGC CATTGCTACA GGCATCGTGG TGTCACGCTC GTCGTTTGGT
 15 ATGGCTTCAT TCAGTCCCG TTCCCAACGA TCAAGGCGAG TTACATGATC CCCATGTTG
 TGCAAAAAG CGGTTAGCTC CTTCCGTCCT CCGATCGTTG TCAGAAGTAA GTTGGCCGCA
 GTGTTATCAC TCATGGTTAT GGCAGCACTG CATAATTCTC TTACTGTCAT GCCATCCGTA
 AGATGCTTTT CTGTGACTGG TGAGTACTCA ACCAAGTCAT TCTGAGAATA GTGTATGCGG
 CGACCGAGTT GCTCTTGCCC GGCGTCAATA CGGGATAATA CCGCGCCACA TAGCAGAACT
 20 TTAAAAGTGC TCATCATTGG AAAACGTTCT TCGGGGCGAA AACTCTCAAG GATCTTACCG
 CTGTTGAGAT CCAGTTCGAT GTAACCCACT CGTGCACCCA ACTGATCTTC AGCATCTTTT
 ACTTTCACCA GCGTTTCTGG GTGAGCAAAA ACAGGAAGGC AAAATGCCGC AAAAAAGGGA
 ATAAGGCGCA CACGGAAATG TTGAATACTC ATACTCTTCC TTTTCAATA TTATTGAAGC
 ATTTATCAGG GTTATTGTCT CATGAGCGGA TACATATTTG AATGTATTTA GAAAAATAAA
 25 CAAATAGGGG TTCCGCGCAC ATTTCCCCGA AAAGTGCCAC CTGACGTCTA AGAAACCATT
 ATTATCATGA CATTAACCTA TAAAAATAGG CGTATCACGA GGCCCTTTCG TC (SEQ ID
 NO:14) .

V1Jneo – Construction of vaccine vector V1Jneo expression vector involved
 removal of the *amp^r* gene and insertion of the *kan^r* gene (neomycin
 30 phosphotransferase). The *amp^r* gene from the pUC backbone of V1J was removed by
 digestion with *SspI* and *Eam1105I* restriction enzymes. The remaining plasmid was
 purified by agarose gel electrophoresis, blunt-ended with T4 DNA polymerase, and
 then treated with calf intestinal alkaline phosphatase. The commercially available
kan^r gene, derived from transposon 903 and contained within the pUC4K plasmid,

was excised using the PstI restriction enzyme, purified by agarose gel electrophoresis, and blunt-ended with T4 DNA polymerase. This fragment was ligated with the VIJ backbone and plasmids with the kan^r gene in either orientation were derived which were designated as VIJneo #'s 1 and 3. Each of these plasmids was confirmed by
 5 restriction enzyme digestion analysis, DNA sequencing of the junction regions, and was shown to produce similar quantities of plasmid as VIJ. Expression of heterologous gene products was also comparable to VIJ for these VIJneo vectors. VIJneo#3, referred to as VIJneo hereafter, was selected which contains the kan^r gene in the same orientation as the amp^r gene in VIJ as the expression construct and
 10 provides resistance to neomycin, kanamycin and G418. The nucleotide sequence of VIJneo is as follows:

```

TCGCGCGTTT CGGTGATGAC GGTGAAAACC TCTGACACAT GCAGCTCCCG GAGACGGTCA
CAGCTTGCTCT GTAAGCGGAT GCCGGGAGCA GACAAGCCCC TCAGGGCGCG TCAGCGGGTG
TTGGCGGGTG TCGGGGCTGG CTTAACTATG CGGCATCAGA GCAGATTGTA CTGAGAGTGC
15 ACCATATGCG GTGTGAAATA CCGCACAGAT GCGTAAGGAG AAAATACCGC ATCAGATTGG
CTATTGGCCA TTGCATACGT TGTATCCATA TCATAATATG TACATTTATA TTGGCTCATG
TCCAACATTA CCGCCATGTT GACATTGATT ATTGACTAGT TATTAATAGT AATCAATTAC
GGGGTCATTA GTTCATAGCC CATATATGGA GTTCCGCGTT ACATAACTTA CGGTAAATGG
CCCGCCTGGC TGACCGCCCA ACGACCCCCG CCCATTGACG TCAATAATGA CGTATGTTCC
20 CATAGTAACG CCAATAGGGA CTTTCCATTG ACGTCAATGG GTGGAGTATT TACGGTAAAC
TGCCCACTTG GCAGTACATC AAGTGTATCA TATGCCAAGT ACGCCCCCTA TTGACGTCAA
TGACGGTAAA TGGCCCGCCT GGCATTATGC CCAGTACATG ACCTTATGGG ACTTTCCTAC
TTGGCAGTAC ATCTACGTAT TAGTCATCGC TATTACCATG GTGATGCGGT TTTGGCAGTA
CATCAATGGG CGTGGATAGC GGTTTGACTC ACGGGGATTT CCAAGTCTCC ACCCCATTGA
25 CGTCAATGGG AGTTTGTTTT GGCACCAAAA TCAACGGGAC TTTCCAAAAT GTCGTAACAA
CTCCGCCCCA TTGACGCAAA TGGGCGGTAG GCGTGTACGG TGGGAGGTCT ATATAAGCAG
AGCTCGTTTA GTGAACCGTC AGATCGCCTG GAGACGCCAT CCACGCTGTT TTGACCTCCA
TAGAAGACAC CGGGACCGAT CCAGCCTCCG CGGCCGGGAA CGGTGCATTG GAACGCGGAT
TCCCCGTGCC AAGAGTGACG TAAGTACCGC CTATAGAGTC TATAGGCCCA CCCCCTTGCG
30 TTCTTATGCA TGCTATACTG TTTTGGCTT GGGGTCTATA CACCCCGCT TCCTCATGTT
ATAGGTGATG GTATAGCTTA GCCTATAGGT GTGGGTATT GACCATTATT GACCACTCCC
CTATTGGTGA CGATACTTTC CATTACTAAT CCATAACATG GCTCTTTGCC ACAACTCTCT
TTATTGGCTA TATGCCAATA CACTGTCCTT CAGAGACTGA CACGGACTCT GTATTTTAC
AGGATGGGGT CTCATTTATT ATTTACAAAT TCACATATAC AACACCACCG TCCCCAGTGC

```


CCGCAGTTTT TATTAAACAT AACGTGGGAT CTCCACGCGA ATCTCGGGTA CGTGTTCCGG
 ACATGGGCTC TTCTCCGGTA GCGGCGGAGC TTCTACATCC GAGCCCTGCT CCCATGCCTC
 CAGCGACTCA TGGTCGCTCG GCAGCTCCTT GCTCCTAACA GTGGAGGCCA GACTTAGGCA
 CAGCACGATG CCCACCACCA CCAGTGTGCC GCACAAGGCC GTGGCGGTAG GGTATGTGTC
 5 TGAAAATGAG CTCGGGGAGC GGGCTTGAC CCGTGACGCA TTTGGAAGAC TTAAGGCAGC
 GGCAGAAGAA GATGCAGGCA GCTGAGTTGT TGTGTTCTGA TAAGAGTCAG AGGTAACCTC
 CGTTGCGGTG CTGTTAACGG TGGAGGGCAG TGTAGTCTGA GCAGTACTCG TTGCTGCCGC
 GCGCGCCACC AGACATAATA GCTGACAGAC TAACAGACTG TTCCTTTCCA TGGGTCTTTT
 CTGCAGTCAC CGTCCTTAGA TCTGCTGTGC CTTCTAGTTG CCAGCCATCT GTTGTGTTGCC
 10 CCTCCCCCGT GCCTTCCTTG ACCCTGGAAG GTGCCACTCC CACTGTCCTT TCCTAATAAA
 ATGAGGAAAT TGCATCGCAT TGTCTGAGTA GGTGTCATTC TATTCTGGGG GGTGGGGTGG
 GGCAGCACAG CAAGGGGGAG GATTGGGAAG ACAATAGCAG GCATGCTGGG GATGCGGTGG
 GCTCTATGGG TACCCAGGTG CTGAAGAATT GACCCGGTTC CTCCTGGGCC AGAAAGAAGC
 AGGCACATCC CCTTCTCTGT GACACACCTT GTCCACGCCC CTGGTTCTTA GTTCCAGCCC
 15 CACTCATAGG AACTCATAG CTCAGGAGGG CTCCGCCTTC AATCCCACCC GCTAAAGTAC
 TTGGAGCGGT CTCTCCCTCC CTCATCAGCC CACCAAACCA AACCTAGCCT CCAAGAGTGG
 GAAGAAATTA AAGCAAGATA GGCTATTAAG TGCAGAGGGA GAGAAAATGC CTCCAACATG
 TGAGGAAGTA ATGAGAGAAA TCATAGAATT TCTTCCGCTT CCTCGCTCAC TGA CTGCTG
 CGCTCGGTCG TTCGGCTGCG GCGAGCGGTA TCAGCTCACT CAAAGGCGGT AATACGGTTA
 20 TCCACAGAAT CAGGGGATAA CGCAGGAAAG AACATGTGAG CAAAAGGCCA GCAAAGGCC
 AGGAACCGTA AAAAGGCCGC GTTGCTGGCG TTTTTCATA GGCTCCGCC CCCTGACGAG
 CATCACAAAA ATCGACGCTC AAGTCAGAGG TGGCGAAACC CGACAGGACT ATAAAGATAC
 CAGGCGTTTC CCCCTGGAAG CTCCCTCGTG CGCTCTCCTG TTCCGACCTT GCCGCTTACC
 GGATACCTGT CCGCCTTTCT CCCTTCGGGA AGCGTGGCGC TTTCTCAATG CTCACGCTGT
 25 AGGTATCTCA GTTCGGTGTA GGTCGTTTCG TCCAAGCTGG GCTGTGTGCA CGAACCCCCC
 GTTCAGCCCC ACCGCTGCGC CTTATCCGGT AACTATCGTC TTGAGTCCAA CCCGGTAAGA
 CACGACTTAT CGCCACTGGC AGCAGCCACT GGTAACAGGA TTAGCAGAGC GAGGTATGTA
 GGCGGTGCTA CAGAGTTCTT GAAGTGGTGG CCTAACTACG GCTACACTAG AAGGACAGTA
 TTTGGTATCT GCGCTCTGCT GAAGCCAGTT ACCTTCGGAA AAAGAGTTGG TAGCTCTTGA
 30 TCCGGCAAAC AAACCACCGC TGGTAGCGGT GGTTTTTTTG TTTGCAAGCA GCAGATTACG
 CGCAGAAAAA AAGGATCTCA AGAAGATCCT TTGATCTTTT CTACGGGGTC TGACGCTCAG
 TGGAACGAAA ACTCACGTTA AGGGATTTTG GTCATGAGAT TATCAAAAAG GATCTTCACC
 TAGATCCTTT TAAATTAAAA ATGAAGTTT AAATCAATCT AAAGTATATA TGAGTAAACT
 TGGTCTGACA GTTACCAATG CTTAATCAGT GAGGCACCTA TCTCAGCGAT CTGTCTATTT

CGTTCATCCA TAGTTGCCTG ACTCCGGGGG GGGGGGGCGC TGAGGTCTGC CTCGTGAAGA
 AGGTGTTGCT GACTCATACC AGGCCTGAAT CGCCCCATCA TCCAGCCAGA AAGTGAGGGA
 GCCACGGTTG ATGAGAGCTT TGTGTAGGT GGACCAGTTG GTGATTTTGA ACTTTTGCTT
 TGCCACGGAA CGGTCTGCGT TGTCGGGAAG ATGCGTGATC TGATCCTTCA ACTCAGCAAA
 5 AGTTCGATTT ATTCAACAAA GCCGCCGTCC CGTCAAGTCA GCGTAATGCT CTGCCAGTGT
 TACAACCAAT TAACCAATTC TGATTAGAAA AACTCATCGA GCATCAAATG AAAC TGCAAT
 TTATTCATAT CAGGATTATC AATACCATAT TTTTGAAAAA GCCGTTTCTG TAATGAAGGA
 GAAAACTCAC CGAGGCAGTT CCATAGGATG GCAAGATCCT GGTATCGGTC TGCGATTCCG
 ACTCGTCCAA CATCAATACA ACCTATTAAT TTCCCCTCGT CAAAAATAAG GTTATCAAGT
 10 GAGAAATCAC CATGAGTGAC GACTGAATCC GGTGAGAATG GCAAAGCTT ATGCATTTCT
 TTCCAGACTT GTTCAACAGG CCAGCCATTA CGCTCGTCAT CAAATCACT CGCATCAACC
 AAACCGTTAT TCATTCGTGA TTGCGCCTGA GCGAGACGAA ATACGCGATC GCTGTTAAAA
 GGACAATTAC AACAGGAAT CGAATGCAAC CGGCGCAGGA AACTGCCAG CGCATCAACA
 ATATTTTCAC CTGAATCAGG ATATTCTTCT AATACCTGGA ATGCTGTTTT CCCGGGGATC
 15 GCAGTGGTGA GTAACCATGC ATCATCAGGA GTACGGATAA AATGCTTGAT GGTGGAAGA
 GGCATAAATT CCGTCAGCCA GTTTAGTCTG ACCATCTCAT CTGTAACATC ATTGGCAACG
 CTACCTTTGC CATGTTTCAG AAACAACTCT GGCGCATCGG GCTTCCATA CAATCGATAG
 ATTGTCGCAC CTGATTGCCC GACATTATCG CGAGCCCAT TATACCCATA TAAATCAGCA
 TCCATGTTGG AATTTAATCG CGGCCTCGAG CAAGACGTTT CCCGTTGAAT ATGGCTCATA
 20 ACACCCCTTG TATTACTGTT TATGTAAGCA GACAGTTTTA TTGTTTCATGA TGATATATTT
 TTATCTGTG CAATGTAACA TCAGAGATTT TGAGACACAA CGTGGCTTTC CCCCCCCCCC
 CATTATTGAA GCATTATCA GGGTTATTGT CTCATGAGCG GATACATATT TGAATGTATT
 TAGAAAAATA AACAAATAGG GGTTCGCGC ACATTTCCCC GAAAAGTGCC ACCTGACGTC
 TAAGAAACCA TTATTATCAT GACATTAACC TATAAAAAATA GGCGTATCAC GAGGCCCTTT
 25 CGTC (SEQ ID NO:15).

VIJns - The expression vector *VIJns* was generated by adding an *SfiI* site to *VIJneo* to facilitate integration studies. A commercially available 13 base pair *SfiI* linker (New England BioLabs) was added at the *KpnI* site within the BGH sequence of the vector. *VIJneo* was linearized with *KpnI*, gel purified, blunted by T4 DNA polymerase, and ligated to the blunt *SfiI* linker. Clonal isolates were chosen by
 30 restriction mapping and verified by sequencing through the linker. The new vector was designated *VIJns*. Expression of heterologous genes in *VIJns* (with *SfiI*) was comparable to expression of the same genes in *VIJneo* (with *KpnI*).

The nucleotide sequence of *VIJns* is as follows:

TCGCGCGTTT CGGTGATGAC GGTGAAAACC TCTGACACAT GCAGCTCCCG GAGACGGTCA
 CAGCTTGTCT GTAAGCGGAT GCCGGGAGCA GACAAGCCCG TCAGGGCGCG TCAGCGGGTG
 TTGGCGGGTG TCGGGGCTGG CTTAACTATG CGGCATCAGA GCAGATTGTA CTGAGAGTGC
 ACCATATGCG GTGTGAAATA CCGCACAGAT GCGTAAGGAG AAAATACCGC ATCAGATTGG
 5 CTATTGGCCA TTGCATACGT TGTATCCATA TCATAATATG TACATTTATA TTGGCTCATG
 TCCAACATTA CCGCCATGTT GACATTGATT ATTGACTAGT TATTAATAGT AATCAATTAC
 GGGGTCATTA GTTCATAGCC CATATATGGA GTTCCGCGTT ACATAACTTA CGGTAAATGG
 CCCGCTGGC TGACCGCCCA ACGACCCCG CCCATTGACG TCAATAATGA CGTATGTTCC
 CATAGTAACG CCAATAGGGA CTTTCCATTG ACGTCAATGG GTGGAGTATT TACGGTAAAC
 10 TGCCCACTTG GCAGTACATC AAGTGTATCA TATGCCAAGT ACGCCCCCTA TTGACGTCAA
 TGACGGTAAA TGGCCCGCCT GGCATTATGC CCAGTACATG ACCTTATGGG ACTTTCCTAC
 TTGGCAGTAC ATCTACGTAT TAGTCATCGC TATTACCATG GTGATGCGGT TTTGGCAGTA
 CATCAATGGG CGTGGATAGC GGTTTGACTC ACGGGGATTT CCAAGTCTCC ACCCCATTGA
 CGTCAATGGG AGTTTGT TTTT GGCACCAAAA TCAACGGGAC TTTCCAAAAT GTCGTAACAA
 15 CTCCGCCCCA TTGACGCAAA TGGGCGGTAG GCGTGTACGG TGGGAGGTCT ATATAAGCAG
 AGCTCGTTTA GTGAACCGTC AGATCGCCTG GAGACGCCAT CCACGCTGTT TTGACCTCCA
 TAGAAGACAC CGGGACCGAT CCAGCCTCCG CGGCCGGGAA CGGTGCATTG GAACGCGGAT
 TCCCCGTGCC AAGAGTGACG TAAGTACCGC CTATAGACTC TATAGGCACA CCCCTTTGGC
 TCTTATGCAT GCTATACTGT TTTTGGCTTG GGGCCTATAC ACCCCCGCTT CCTTATGCTA
 20 TAGGTGATGG TATAGCTTAG CCTATAGGTG TGGGTTATTG ACCATTATTG ACCACTCCCC
 TATTGGTGAC GATACTTTCC ATTACTAATC CATAACATGG CTCTTTGCCA CAACTATCTC
 TATTGGCTAT ATGCCAATAC TCTGTCTTC AGAGACTGAC ACGGACTCTG TATTTTTTACA
 GGATGGGGTC CCATTATTAT TTTACAAATT CACATATACA ACAACGCCGT CCCCCGTGCC
 CGCAGTTTTT ATTAACATA GCGTGGGATC TCCACGCGAA TCTCGGGTAC GTGTTCCGGA
 25 CATGGGCTCT TCTCCGGTAG CGGCGGAGCT TCCACATCCG AGCCCTGGTC CCATGCCTCC
 AGCGGCTCAT GGTCGCTCGG CAGCTCCTTG CTCCTAACAG TGGAGGCCAG ACTTAGGCAC
 AGACAATGC CCACCACCAC CAGTGTGCCG CACAAGGCCG TGGCGGTAGG GTATGTGTCT
 GAAAATGAGC GTGGAGATTG GGCTCGCACG GCTGACGCAG ATGGAAGACT TAAGGCAGCG
 GCAGAAGAAG ATGCAGGCAG CTGAGTTGTT GTATTCTGAT AAGAGTCAGA GGTAACTCCC
 30 GTTGCGGTGC TGTTAACGGT GGAGGGCAGT GTAGTCTGAG CAGTACTCGT TGCTGCCGCG
 CGCGCCACCA GACATAATAG CTGACAGACT AACAGACTGT TCCTTTCCAT GGGTCTTTTC
 TGCAGTCACC GTCCTTAGAT CTGCTGTGCC TTCTAGTTGC CAGCCATCTG TTGTTTGCCC
 CTCCCCGTG CCTTCCTTGA CCCTGGAAGG TGCCACTCCC ACTGTCTTTT CCTAATAAAA
 TGAGGAAATT GCATCGCATT GTCTGAGTAG GTGTCATTCT ATTCTGGGGG GTGGGGTGGG

GCAGGACAGC AAGGGGGAGG ATTGGGAAGA CAATAGCAGG CATGCTGGGG ATGCGGTGGG
 CTCTATGGCC GCTGCGGCCA GGTGCTGAAG AATTGACCCG GTTCCTCCTG GGCCAGAAAG
 AAGCAGGCAC ATCCCCTTCT CTGTGACACA CCCTGTCCAC GCCCCTGGTT CTTAGTTCCA
 GCCCCACTCA TAGGACACTC ATAGCTCAGG AGGGCTCCGC CTTCAATCCC ACCCGCTAAA
 5 GTACTTGGAG CGGTCTCTCC CTCCCTCATC AGCCCACCAA ACCAAACCTA GCCTCCAAGA
 GTGGGAAGAA ATTAAAGCAA GATAGGCTAT TAAGTGCAGA GGGAGAGAAA ATGCCTCCAA
 CATGTGAGGA AGTAATGAGA GAAATCATAG AATTTCTTCC GCTTCCTCGC TCACTGACTC
 GCTGCGCTCG GTCGTTCCGC TGCGGCGAGC GGTATCAGCT CACTCAAAGG CGGTAATACG
 GTTATCCACA GAATCAGGGG ATAACGCAGG AAAGAACATG TGAGCAAAAG GCCAGCAAAA
 10 GGCCAGGAAC CGTAAAAAGG CCGCGTTGCT GGCGTTTTTC CATAGGCTCC GCCCCCTGA
 CGAGCATCAC AAAAATCGAC GCTCAAGTCA GAGGTGGCGA AACCCGACAG GACTATAAAG
 ATACCAGGCG TTTCCCTCTG GAAGCTCCCT CGTGCGCTCT CCTGTTCCGA CCCTGCCGCT
 TACCGGATAC CTGTCCGCCT TTCTCCCTTC GGAAGCGTG GCGCTTTCTC ATAGCTCACG
 CTGTAGGTAT CTCAGTTCGG TGTAGGTCGT TCGCTCCAAG CTGGGCTGTG TGCACGAACC
 15 CCCCCTTCAG CCCGACCGCT GCGCCTTATC CGGTAACATAT CGTCTTGAGT CCAACCCGGT
 AAGACACGAC TTATCGCCAC TGGCAGCAGC CACTGGTAAC AGGATTAGCA GAGCGAGSTA
 TGTAGGCGGT GCTACAGAGT TCTTGAAGTG GTGGCCTAAC TACGGCTACA CTAGAAGAAC
 AGTATTTGGT ATCTGCGCTC TGCTGAAGCC AGTTACCTTC GGAAAAAGAG TTGGTAGCTC
 TTGATCCGGC AAACAAACCA CCGCTGGTAG CCGTGTTTTC TTTGTTTGCA AGCAGCAGAT
 20 TACGCGCAGA AAAAAAGGAT CTCAAGAAGA TCCTTTGATC TTTTCTACGG GGTCTGACGC
 TCAGTGGAAC GAAACTCAC GTTAAGGGAT TTTGGTCATG AGATTATCAA AAAGGATCTT
 CACCTAGATC CTTTTAAATT AAAAATGAAG TTTTAAATCA ATCTAAAGTA TATATGAGTA
 AACTTGGTCT GACAGTTACC AATGCTTAAT CAGTGAGGCA CCTATCTCAG CGATCTGTCT
 ATTTCTGTTCA TCCATAGTTG CCTGACTCGG GGGGGGGGGG CGCTGAGGTC TGCCTCGTGA
 25 AGAAGGTGTT GCTGACTCAT ACCAGGCCTG AATCGCCCCA TCATCCAGCC AGAAAGTGAG
 GGAGCCACGG TTGATGAGAG CTTTGTGTGA GGTGGACCAG TTGGTGATTT TGAACTTTTC
 CTTTGCCACG GAACGGTCTG CGTTGTCGGG AAGATGCGTG ATCTGATCCT TCAACTCAGC
 AAAAGTTCGA TTTATTCAAC AAAGCCGCCG TCCCGTCAAG TCAGCGTAAT GCTCTGCCAG
 TGTTACAACC AATTAACCAA TTCTGATTAG AAAAATCAT CGAGCATCAA ATGAAACTGC
 30 AATTTATTCA TATCAGGATT ATCAATACCA TATTTTGTAA AAAGCCGTTT CTGTAATGAA
 GGAGAAAACT CACCGAGGCA GTTCCATAGG ATGGCAAGAT CCTGGTATCG GTCTGCGATT
 CCGACTCGTC CAACATCAAT ACAACCTATT AATTTCCCTT CGTCAAAAAT AAGGTTATCA
 AGTGAGAAAT CACCATGAGT GACGACTGAA TCCGGTGAGA ATGGCAAAAG CTTATGCATT
 TCTTTCCAGA CTTGTTCAAC AGGCCAGCCA TTACGCTCGT CATCAAAATC ACTCGCATCA

ACCAAACCGT TATTCATTCTG TGATTGCGCC TGAGCGAGAC GAAATACGCG ATCGCTGTTA
 AAAGGACAAT TACAAACAGG AATCGAATGC AACCGGCGCA GGAACACTGC CAGCGCATCA
 ACAATATTTT CACCTGAATC AGGATATTCT TCTAATACCT GGAATGCTGT TTTCCCGGGG
 ATCGCAGTGG TGAGTAACCA TGCATCATCA GGAGTACGGA TAAATGCTT GATGGTCGGA
 5 AGAGGCATAA ATTCCGTCAG CCAGTTTAGT CTGACCATCT CATCTGTAAC ATCATTGGCA
 ACGCTACCTT TGCCATGTTT CAGAAACAAC TCTGGCGCAT CGGGCTTCCC ATACAATCGA
 TAGATTGTCG CACCTGATTG CCCGACATTA TCGCGAGCCC ATTTATACCC ATATAAATCA
 GCATCCATGT TGGAATTTAA TCGCGGCCTC GAGCAAGACG TTTCCCGTTG AATATGGCTC
 ATAACACCCC TTGTATTACT GTTTATGTAA GCAGACAGTT TTATTGTTCA TGATGATATA
 10 TTTTATCTT GTGCAATGTA ACATCAGAGA TTTTGAGACA CAACGTGGCT TTCCCCCCCC
 CCCCATTTAT GAAGCATTTA TCAGGGTTAT TGTCTCATGA GCGGATACAT ATTTGAATGT
 ATTTAGAAAA ATAAACAAAT AGGGGTTCG CGCACATTTC CCCGAAAAGT GCCACCTGAC
 GTCTAAGAAA CCATTATTAT CATGACATTA ACCTATAAAA ATAGGCGTAT CACGAGGCCC
 TTTCGTC (SEQ ID NO:16).

15 The underlined nucleotides of SEQ ID NO:16 represent the SfiI site introduced into the Kpn I site of VIJneo.

VIJns-tPA – The vaccine vector VIJns-tPA was constructed in order to fuse an heterologous leader peptide sequence to the nef DNA constructs of the present invention. More specifically, the vaccine vector VIJns was modified to include the

20 human tissue-specific plasminogen activator (tPA) leader. As an exemplification, but by no means a limitation of generating a nef DNA construct comprising an amino-terminal leader sequence, plasmid VIJneo was modified to include the human tissue-specific plasminogen activator (tPA) leader. Two synthetic complementary oligomers were annealed and then ligated into VIJneo which had been BglII digested. The

25 sense and antisense oligomers were 5' GATCACCATGGATGCAATGAAGAGAG GGCTCTGCTGTGTGCTGCTGCTGTGTGGAGCAGTCTTCGTTTCGCCCAG CGA-3' (SEQ ID NO:17); and, 5'-GATCTCGCTGGGCGAAACGAAGACTGC TCCACACAGCAGCAGCACACAGCAGAGCCCTCTCTTCATTGCATCCAT GGT-3' (SEQ ID NO:18). The Kozak sequence is underlined in the sense oligomer.

30 These oligomers have overhanging bases compatible for ligation to BglII-cleaved sequences. After ligation the upstream BglII site is destroyed while the downstream BglII is retained for subsequent ligations. Both the junction sites as well as the entire tPA leader sequence were verified by DNA sequencing. Additionally, in order to conform with VIJns (=VIJneo with an SfiI site), an SfiI restriction site was placed at

the KpnI site within the BGH terminator region of V1Jneo-tPA by blunting the KpnI site with T4 DNA polymerase followed by ligation with an SfiI linker (catalogue #1138, New England Biolabs), resulting in V1Jns-tPA. This modification was verified by restriction digestion and agarose gel electrophoresis.

5 The V1Jns-tpa vector nucleotide sequence is as follows:

```

TCGCGCGTTT CCGTGATGAC GGTGAAAACC TCTGACACAT GCAGTCCCG GAGACGGTCA
CAGCTTGTCT GTAAGCGGAT GCCGGGAGCA GACAAGCCCG TCAGGGCGCG TCAGCGGGTG
TTGGCGGGTG TCGGGGCTGG CTTAACTATG CGGCATCAGA GCAGATTGTA CTGAGAGTGC
ACCATATGCG GTGTGAAATA CCGCACAGAT GCGTAAGGAG AAAATACCGC ATCAGATTGG
10 CTATTGGCCA TTGCATACGT TGTATCCATA TCATAATATG TACATTTATA TTGGCTCATG
TCCAACATTA CCGCCATGTT GACATTGATT ATTGACTAGT TATTAATAGT AATCAATTAC
GGGGTCATTA GTTCATAGCC CATATATGGA GTTCCGCGTT ACATAACTTA CGGTAAATGG
CCCGCCTGGC TGACCGCCCA ACGACCCCG CCCATTGACG TCAATAATGA CGTATGTTCC
CATAGTAACG CCAATAGGGA CTTTCCATTG ACGTCAATGG GTGGAGTATT TACGGTAAAC
15 TGCCCCCTTG GCAGTACATC AAGTGTATCA TATGCCAAGT ACGCCCCCTA TTGACGTCAA
TGACGGTAAA TGGCCCGCCT GGCATTATGC CCAGTACATG ACCTTATGGG ACTTTCCTAC
TTGGCAGTAC ATCTACGTAT TAGTCATCGC TATTACCATG GTGATGCGGT TTTGGCAGTA
CATCAATGGG CGTGGATAGC GGTTTGACTC ACGGGGATTT CCAAGTCTCC ACCCCATTGA
CGTCAATGGG AGTTTGTTTT GGCACCAAAA TCAACGGGAC TTTCCAAAAT GTCGTAACAA
20 CTCCGCCCCA TTGACGCAA TGGCGGGTAG GCGTGACGG TGGGAGGTCT ATATAAGCAG
AGCTCGTTTA GTGAACCGTC AGATCGCCTG GAGACGCCAT CCACGCTGTT TTGACCTCCA
TAGAAGACAC CGGGACCGAT CCAGCCTCCG CGGCCGGGAA CGGTGCATTG GAACGCGGAT
TCCCCGTGCC AAGAGTGACG TAAGTACCGC CTATAGACTC TATAGGCACA CCCCTTTGGC
TCTTATGCAT GCTATACTGT TTTTGGCTTG GGGCCTATAC ACCCCGCTT CTTATGCTA
25 TAGGTGATGG TATAGCTTAG CCTATAGGTG TGGGTATTG ACCATTATTG ACCACTCCCC
TATTGGTGAC GATACTTTCC ATTACTAATC CATAACATGG CTCTTTGCCA CAACTATCTC
TATTGGCTAT ATGCCAATAC TCTGTCTTTC AGAGACTGAC ACGGACTCTG TATTTTACA
GGATGGGGTC CCATTTATTA TTTACAAATT CACATATACA ACAACGCCGT CCCCCGTGCC
CGCAGTTTTT ATTAAACATA GCGTGGGATC TCCACGCGAA TCTCGGGTAC GTGTCCGGA
30 CATGGGCTCT TCTCCGGTAG CGGCGGAGCT TCCACATCCG AGCCCTGGTC CCATGCCTCC
AGCGGCTCAT GGTCGCTCGG CAGCTCCTTG CTCTAACAG TGGAGGCCAG ACTTAGGCAC
AGCACAATGC CCACCACCAC CAGTGTGCCG CACAAGGCCG TGGCGGTAGG GTATGTGTCT
GAAAATGAGC GTGGAGATTG GGCTCGCAG GCTGACGCAG ATGGAAGACT TAAGGCAGCG
GCAGAAGAAG ATGCAGGCAG CTGAGTTGTT GTATTCTGAT AAGAGTCAGA GGTAACCCC

```

GTTGCGGTGC TGTAAACGGT GGAGGGCAGT GTAGTCTGAG CAGTACTCGT TGCTGCCGCG
 CGCGCCACCA GACATAATAG CTGACAGACT AACAGACTGT TCCTTTCCAT GGGTCTTTTC
 TGCAGTCACC GTCCTTAGAT CACCATGGAT GCAATGAAGA GAGGGCTCTG CTGTGTGCTG
CTGCTGTGTG GAGCAGTCTT CGTTTCGCCC AGCGAGATCT GCTGTGCCTT CTAGTTGCCA
 5 GCCATCTGTT GTTTGCCCCCT CCCCCGTGCC TTCCTTGACC CTGGAAGGTG CCACTCCCAC
 TGTCCTTTCC TAATAAAATG AGGAAATTGC ATCGCATTGT CTGAGTAGGT GTCATTCTAT
 TCTGGGGGGT GGGGTGGGGC AGGACAGCAA GGGGGAGGAT TGGGAAGACA ATAGCAGGCA
 TGCTGGGGAT GCGGTGGGCT CTATGGCCGC TGCGGCCAGG TGCTGAAGAA TTGACCCGGT
 TCCTCCTGGG CCAGAAAGAA GCAGGCACAT CCCCTTCTCT GTGACACACC CTGTCCACGC
 10 CCCTGGTTCT TAGTTCCAGC CCCACTCATA GGACACTCAT AGCTCAGGAG GGCTCCGCCT
 TCAATCCCAC CCGCTAAAGT ACTTGGAGCG GTCTCTCCCT CCCTCATCAG CCCACCAAAC
 CAAACCTAGC CTCCAAGAGT GGGAAGAAAT TAAAGCAAGA TAGGCTATTA AGTGCAGAGG
 GAGAGAAAAT GCCTCCAACA TGTGAGGAAG TAATGAGAGA AATCATAGAA TTTCTTCCGC
 TTCCTCGCTC ACTGACTCGC TGCCTCGGT CGTTCGGCTG CGGCGAGCGG TATCAGCTCA
 15 CTCAAAGGCG GTAATACGGT TATCCACAGA ATCAGGGGAT AACGCAGGAA AGAACATGTG
 AGCAAAAGGC CAGCAAAAGG CCAGGAACCG TAAAAAGGCC GCGTTGCTGG CGTTTTTCCA
 TAGGCTCCGC CCCCTGACG AGCATCACAA AAATCGACGC TCAAGTCAGA GGTGGCGAAA
 CCCGACAGGA CTATAAAGAT ACCAGGCGTT TCCCCCTGGA AGCTCCCTCG TGCCTCTCC
 TGTTCGACC CTGCCGCTTA CCGGATACCT GTCCGCCTTT CTCCCTTCGG GAAGCGTGGC
 20 GCTTTCTCAT AGCTCACGCT GTAGGTATCT CAGTTCGGTG TAGGTCGTTT GCTCCAAGCT
 GGGCTGTGTG CACGAACCCC CCGTTCAGCC CGACCGCTGC GCCTTATCCG GTAACATCG
 TCTTAGAGTCC AACCCGGTAA GACACGACTT ATCGCCACTG GCAGCAGCCA CTGGTAACAG
 GATTAGCAGA GCGAGGTATG TAGGCGGTGC TACAGAGTTC TTGAAGTGGT GGCCTAACTA
 CGGCTACACT AGAAGAACAG TATTTGGTAT CTGCGCTCTG CTGAAGCCAG TTACCTTCGG
 25 AAAAAGAGTT GGTAGCTCTT GATCCGGCAA ACAAACCACC GCTGGTAGCG GTGGTTTTTT
 TGTTTGCAAG CAGCAGATTA CGCGCAGAAA AAAAGGATCT CAAGAAGATC CTTTGATCTT
 TTCTACGGGG TCTGACGCTC AGTGAACGA AACTCACGT TAAGGGATTT TGGTCATGAG
 ATTATCAAAA AGGATCTTCA CCTAGATCCT TTTAAATTAA AAATGAAGTT TTAAATCAAT
 CTAAAGTATA TATGAGTAAA CTTGGTCTGA CAGTTACCAA TGCTTAATCA GTGAGGCACC
 30 TATCTCAGCG ATCTGTCTAT TTCGTTATC CATAGTTGCC TGAATCGGGG GGGGGGGGCG
 CTGAGGTCTG CCTCGTGAAG AAGGTGTTGC TGACTCATAC CAGGCCTGAA TCGCCCCATC
 ATCCAGCCAG AAAGTGAGGG AGCCACGGTT GATGAGAGCT TTGTTGTAGG TGGACCAGTT
 GGTGATTTTG AACTTTTGCT TTGCCACGGA ACGGTCTGCG TTGTCGGGAA GATGCGTGAT
 CTGATCCTTC AACTCAGCAA AAGTTCGATT TATTCAACAA AGCCGCCGTC CCGTCAAGTC

AGCGTAATGC TCTGCCAGTG TTACAACCAA TTAACCAATT CTGATTAGAA AAACCTCATCG
 AGCATCAAAT GAAACTGCAA TTTATTTCATA TCAGGATTAT CAATACCATA TTTTGGAAAA
 AGCCGTTTCT GTAATGAAGG AGAAACTCA CCGAGGCAGT TCCATAGGAT GGCAAGATCC
 TGGTATCGGT CTGCGATTCC GACTCGTCCA ACATCAATAC AACCTATTAA TTTCCCCTCG
 5 TCAAAAATAA GGTTATCAAG TGAGAAATCA CCATGAGTGA CGACTGAATC CGGTGAGAAT
 GGCAAAAGCT TATGCATTTC TTTCCAGACT TGTTCAACAG GCCAGCCATT ACGCTCGTCA
 TCAAAATCAC TCGCATCAAC CAAACCGTTA TTCATTCTGT ATTGCGCCTG AGCGAGACGA
 AATACGCGAT CGCTGTAAA AGGACAATTA CAAACAGGAA TCGAATGCAA CCGGCGCAGG
 AACACTGCCA GCGCATCAAC AATATTTTCA CCTGAATCAG GATATTCTTC TAATACCTGG
 10 AATGCTGTTT TCCCGGGGAT CGCAGTGGTG AGTAACCATG CATCATCAGG AGTACGGATA
 AAATGCTTGA TGGTCGGAAG AGGCATAAAT TCCGTCAGCC AGTTTAGTCT GACCATCTCA
 TCTGTAACAT CATTGGCAAC GCTACCTTTG CCATGTTTCA GAAACAATC TGGCGCATCG
 GGCTTCCCAT ACAATCGATA GATTGTCGCA CCTGATTGCC CGACATTATC GCGAGCCCAT
 TTATACCCAT ATAAATCAGC ATCCATGTTG GAATTTAATC GCGGCCTCGA GCAAGACGTT
 15 TCCCGTTGAA TATGGTCAT AACACCCCTT GTATTACTGT TTATGTAAGC AGACAGTTTT
 ATTGTTTCATG ATGATATATT TTTATCTTGT GCAATGTAAC ATCAGAGATT TTGAGACACA
 ACGTGGCTTT CCCCCCCCCC CCATTATTGA AGCATTTATC AGGGTTATTG TCTCATGAGC
 GGATACATAT TTGAATGTAT TTAGAAAAAT AAACAAATAG GGGTTCCGCG CACATTTCCC
 CGAAAAGTGC CACCTGACGT CTAAGAAACC ATTATTATCA TGACATTAAC CTATAAAAAAT
 20 AGGCGTATCA CGAGGCCCTT TCGTC (SEQ ID NO:9) .

The underlined nucleotides of SEQ ID NO:9 represent the SfiI site introduced into the KpnI site of VIJneo while the underlined/italicized nucleotides represent the human tPA leader sequence.

25 *VIR* – Vaccine vector *VIR* was constructed to obtain a minimum-sized vaccine vector without unneeded DNA sequences, which still retained the overall optimized heterologous gene expression characteristics and high plasmid yields that VIJ and VIJns afford. It was determined that (1) regions within the pUC backbone comprising the *E. coli* origin of replication could be removed without affecting plasmid yield from bacteria; (2) the 3'-region of the *kan^r* gene following the
 30 kanamycin open reading frame could be removed if a bacterial terminator was inserted in its place; and, (3) ~300 bp from the 3'- half of the BGH terminator could be removed without affecting its regulatory function (following the original KpnI restriction enzyme site within the BGH element). *VIR* was constructed by using PCR to synthesize three segments of DNA from VIJns representing the CMVintA

promoter/BGH terminator, origin of replication, and kanamycin resistance elements, respectively. Restriction enzymes unique for each segment were added to each segment end using the PCR oligomers: SspI and XhoI for CMVintA/BGH; EcoRV and BamHI for the *kan*^r gene; and, BclI and SalI for the *ori*^r. These enzyme sites were chosen because they allow directional ligation of each of the PCR-derived DNA segments with subsequent loss of each site: EcoRV and SspI leave blunt-ended DNAs which are compatible for ligation while BamHI and BclI leave complementary overhangs as do SalI and XhoI. After obtaining these segments by PCR each segment was digested with the appropriate restriction enzymes indicated above and then ligated together in a single reaction mixture containing all three DNA segments. The 5'-end of the *ori*^r was designed to include the T2 rho independent terminator sequence that is normally found in this region so that it could provide termination information for the kanamycin resistance gene. The ligated product was confirmed by restriction enzyme digestion (>8 enzymes) as well as by DNA sequencing of the ligation junctions. DNA plasmid yields and heterologous expression using viral genes within VIR appear similar to VIJns. The net reduction in vector size achieved was 1346 bp (VIJns = 4.86 kb; VIR = 3.52 kb). PCR oligomer sequences used to synthesize VIR (restriction enzyme sites are underlined and identified in brackets following sequence) are as follows: (1) 5'-GGTACAAATATGGCTATTGGC CATTGCATACG-3' (SEQ ID NO:20) [SspI]; (2) 5'-CCACATCTCGAGGAA CCGGGTCAATTCTTCAGCACC-3' (SEQ ID NO:21) [XhoI] (for CMVintA/BGH segment); (3) 5'-GGTACAGATATCGGAAAGCCACGTTGTG TCTCAAATC-3' (SEQ ID NO:22) [EcoRV]; (4) 5'-CACATGGATCCGTAATGCTCTGCCAGTGT TACAACC-3' (SEQ ID NO:23) [BamHI], (for kanamycin resistance gene segment) (5) 5'-GGTACATG ATCACGTAGAAAAGATCAAAGGATCTTCTTG-3' (SEQ ID NO:24) [BclI]; (6) 5'-CCACATGTCGACCCGTAAAAAGGCCGCGTTGCTGG-3' (SEQ ID NO:25): [SalI], (for *E. coli* origin of replication).

The nucleotide sequence of vector VIR is as follows:

TCCGCGGTTT CGGTGATGAC GGTGAAAACC TCTGACACAT GCAGTCCCG GAGACGGTCA
 CAGCTTGTCT GTAAGCGGAT GCCGGGAGCA GACAAGCCCG TCAGGGCGCG TCAGCGGGTG
 TTGGCGGGTG TCGGGGCTGG CTTAACTATG CGGCATCAGA GCAGATTGTA CTGAGAGTGC
 ACCATATGCG GTGTGAAATA CCGCACAGAT GCGTAAGGAG AAAATACCGC ATCAGATTGG
 CTATTGGCCA TTGCATACGT TGTATCCATA TCATAATATG TACATTATA TTGGCTCATG
 TCCAACATTA CCGCCATGTT GACATTGATT ATTGACTAGT TATTAATAGT AATCAATTAC

GGGGTCATTA GTTCATAGCC CATATATGGA GTTCCGCGTT ACATAACTTA CGGTAAATGG
 CCCGCCTGGC TGACCGCCCA ACGACCCCGG CCCATTGACG TCAATAATGA CGTATGTTCC
 CATAGTAACG CCAATAGGGA CTTTCCATTG ACGTCAATGG GTGGAGTATT TACGGTAAAC
 TGCCCACTTG GCAGTACATC AAGTGATCA TATGCCAAGT ACGCCCCCTA TTGACGTCAA
 5 TGACGGTAAA TGGCCCGCCT GGCATTATGC CCAGTACATG ACCTTATGGG ACTTTCCTAC
 TTGGCAGTAC ATCTACGTAT TAGTCATCGC TATTACCATG GTGATGCGGT TTTGGCAGTA
 CATCAATGGG CGTGGATAGC GGTGTGACTC ACGGGGATTT CCAAGTCTCC ACCCCATTGA
 CGTCAATGGG AGTTTGT TTTT GGCACCAAAA TCAACGGGAC TTTCCAAAAT GTCGTAACAA
 CTCCGCCCCA TTGACGCAAA TGGGCGGTAG GCGTGACGG TGGGAGGTCT ATATAAGCAG
 10 AGCTCGTTTA GTGAACCGTC AGATCGCCTG GAGACGCCAT CCACGCTGTT TTGACCTCCA
 TAGAAGACAC CGGGACCGAT CCAGCCTCCG CGGCCGGGAA CGGTGCATTG GAACGCGGAT
 TCCCCGTGCC AAGAGTGACG TAAGTACCGC CTATAGAGTC TATAGGCCCA CCCCCTTGGC
 TTCTTATGCA TGCTATACTG TTTTGGCTT GGGGTCTATA CACCCCGCT TCCTCATGTT
 ATAGGTGATG GTATAGCTTA GCCTATAGGT GTGGGTATT GACCATTATT GACCACTCCC
 15 CTATTGGTGA CGATACTTTC CATTACTAAT CCATAACATG GCTCTTTGCC ACAACTCTCT
 TTATTGGCTA TATGCCAATA CACTGTCTT CAGAGACTGA CACGGACTCT GTATTTTTAC
 AGGATGGGGT CTCATTTATT ATTTACAAAT TCACATATAC AACACCACCG TCCCCAGTGC
 CCGCAGTTT TATTAAACAT AACGTGGGAT CTCCACGCGA ATCTCGGGTA CGTGTTCGGG
 ACATGGGCTC TTCTCCGGTA GCGGCGGAGC TTCTACATCC GAGCCCTGCT CCCATGCCTC
 20 CAGCGACTCA TGGTCGCTCG GCAGCTCCTT GCTCCTAACA GTGGAGGCCA GACTTAGGCA
 CAGCAGATG CCCACCACCA CCAGTGTGCC GCACAAGGCC GTGGCGGTAG GGTATGTGTC
 TGAAAATGAG CTCGGGGAGC GGGCTTGAC CGCTGACGCA TTTGGAAGAC TTAAGGCAGC
 GGCAGAAGAA GATGCAGGCA GCTGAGTTGT TGTGTTCTGA TAAGAGTCAG AGGTAACCTC
 CGTTGCGGTG CTGTTAACGG TGGAGGGCAG TGTAGTCTGA GCAGTACTCG TTGCTGCCGC
 25 GCGCGCCACC AGACATAATA GCTGACAGAC TAACAGACTG TTCCTTTCCA TGGGTCTTTT
 CTGCAGTCAC CGTCCTTAGA TCTGCTGTGC CTTCTAGTTG CCAGCCATCT GTTGTTTGCC
 CCTCCCCCGT GCCTTCCTTG ACCCTGGAAG GTGCCACTCC CACTGTCTT TCCTAATAAA
 ATGAGGAAAT TGCATCGCAT TGTCTGAGTA GGTGTCATTC TATTCTGGGG GGTGGGGTGG
 GGCAGCACAG CAAGGGGGAG GATTGGGAAG ACAATAGCAG GCATGCTGGG GATGCGGTGG
 30 GCTCTATGGG TACCCAGGTG CTGAAGAATT GACCCGGTTC CTCCTGGGCC AGAAAGAAGC
 AGGCACATCC CTTCTCTGT GACACACCT GTCCACGCCC CTGGTTCTTA GTTCCAGCCC
 CACTCATAGG AACTCATAG CTCAGGAGGG CTCCGCCTTC AATCCCACCC GCTAAAGTAC
 TTGGAGCGGT CTCTCCCTCC CTCATCAGCC CACCAAACCA AACCTAGCCT CCAAGAGTGG
 GAAGAAATTA AAGCAAGATA GGCTATTAAG TGCAGAGGGA GAGAAAATGC CTCCAACATG

TGAGGAAGTA ATGAGAGAAA TCATAGAATT TCTTCCGCTT CCTCGCTCAC TGA CTGCTG
 CGCTCGGTTCG TTCGGCTGCG GCGAGCGGTA TCAGCTCACT CAAAGGCGGT AATACGGTTA
 TCCACAGAAT CAGGGGATAA CGCAGGAAAG AACATGTGAG CAAAAGGCCA GCAAAAGGCC
 AGGAACCGTA AAAAGGCCGC GTTGCTGGCG TTTTTCATA GGCTCCGCCC CCCTGACGAG
 5 CATCACAAAA ATCGACGCTC AAGTCAGAGG TGGCGAAACC CGACAGGACT ATAAAGATAC
 CAGGCGTTTC CCCCTGGAAG CTCCCTCGTG CGCTCTCCTG TTCCGACCCT GCCGCTTACC
 GGATACCTGT CCGCCTTTCT CCCTTCGGGA AGCGTGCGC TTTCTCAATG CTCACGCTGT
 AGGTATCTCA GTTCGGTGTA GGTCGTTGCG TCCAAGCTGG GCTGTGTGCA CGAACCCCCC
 GTTCAGCCCG ACCGCTGCGC CTTATCCGGT AACTATCGTC TTGAGTCCAA CCCGGTAAGA
 10 CACGACTTAT CGCCACTGGC AGCAGCCACT GGTAACAGGA TTAGCAGAGC GAGGTATGTA
 GGCGGTGCTA CAGAGTTCTT GAAGTGGTGG CCTAACTACG GCTACACTAG AAGGACAGTA
 TTTGGTATCT GCGCTCTGCT GAAGCCAGTT ACCTTCGGAA AAAGAGTTGG TAGCTCTTGA
 TCCGGCAAAC AAACCACCGC TGGTAGCGGT GGTTTTTTTG TTTGCAAGCA GCAGATTACG
 CGCAGAAAAA AAGGATCTCA AGAAGATCCT TTGATCTTTT CTACGGGGTC TGACGCTCAG
 15 TGGAACGAAA ACTCACGTTA AGGGATTTTG GTCATGAGAT TATCAAAAAG GATCTTCACC
 TAGATCCTTT TAAATTAAAA ATGAAGTTTT AAATCAATCT AAAGTATATA TGAGTAAACT
 TGGTCTGACA GTTACCAATG CTTAATCAGT GAGGCACCTA TCTCAGCGAT CTGTCTATTT
 CGTTCATCCA TAGTTGCCTG ACTCCGGGGG GGGGGGGCGC TGAGGTCTGC CTCGTGAAGA
 AGGTGTTGCT GACTCATACC AGGCCTGAAT CGCCCCATCA TCCAGCCAGA AAGTGAGGGA
 20 GCCACGGTTG ATGAGAGCTT TGTGTAGGT GGACCAGTTG GTGATTTTGA ACTTTTGCTT
 TGCCACGGAA CGGTCTGCGT TGTCCGGAAG ATGCGTGATC TGATCCTTCA ACTCAGCAAA
 AGTTCGATTT ATTCAACAAA GCCGCCGTCC CGTCAAGTCA GCGTAATGCT CTGCCAGTGT
 TACAACCAAT TAACCAATTC TGATTAGAAA AACTCATCGA GCATCAAATG AAAC TGCAAT
 TTATTCATAT CAGGATTATC AATACCATAT TTTTGAAAAA GCCGTTTCTG TAATGAAGGA
 25 GAAAACTCAC CGAGGCAGTT CCATAGGATG GCAAGATCCT GGTATCGGTC TGCGATTCCG
 ACTCGTCCAA CATCAATACA ACCTATTAAT TTCCCCTCGT CAAAATAAG GTTATCAAGT
 GAGAAATCAC CATGAGTGAC GACTGAATCC GGTGAGAATG GCAAAAGCTT ATGCATTTCT
 TTCCAGACTT GTTCAACAGG CCAGCCATTA CGCTCGTCAT CAAAATCACT CGCATCAACC
 AAACCGTTAT TCATTGCTGA TTGCGCCTGA GCGAGACGAA ATACGCGATC GCTGTTAAAA
 30 GGACAATTAC AAACAGGAAT CGAATGCAAC CGGCGCAGGA AACTGCCAG CGCATCAACA
 ATATTTTCAC CTGAATCAGG ATATTCTTCT AATACCTGGA ATGCTGTTTT CCCGGGGATC
 GCAGTGGTGA GTAACCATGC ATCATCAGGA GTACGGATAA AATGCTTGAT GGTGGAAGA
 GGCATAAATT CCGTCAGCCA GTTTAGTCTG ACCATCTCAT CTGTAACATC ATTGGCAACG
 CTACCTTTGC CATGTTTCAG AAACAACCTCT GGCGCATCGG GCTTCCCATA CAATCGATAG

ATTGTCGCAC CTGATTGCCC GACATTATCG CGAGCCCATT TATACCCATA TAAATCAGCA
 TCCATGTTGG AATTTAATCG CGGCCTCGAG CAAGACGTTT CCCGTTGAAT ATGGCTCATA
 ACACCCCTTG TATTACTGTT TATGTAAGCA GACAGTTTTA TTGTTTCATGA TGATATATTT
 TTATCTTGTT CAATGTAACA TCAGAGATTT TGAGACACAA CGTGGCTTTC CCCCCCCCCC
 5 CATTATTGAA GCATTTATCA GGGTTATTGT CTCATGAGCG GATACATATT TGAATGTATT
 TAGAAAAATA AACAAATAGG GGTTCGCGC ACATTTCCCC GAAAAGTGCC ACCTGACGTC
 TAAGAAACCA TTATTATCAT GACATTAACC TATAAAAATA GCGGTATCAC GAGGCCCTTT
 CGTC (SEQ ID NO:26).

10

EXAMPLE 2

Codon Optimized HIV-1 Nef and HIV-1 Nef Derivatives as DNA Vector Vaccines

HIV-1 Nef Vaccine Vectors - Codon optimized nef gene coding for wt Nef protein of HIV-1 jrf1 isolate was assembled from complementary, overlapping synthetic oligonucleotides by polymerase chain reaction (PCR). The PCR primers
 15 used were designed in such that a BglII site was included in the extension of 5' primer and an SrfI site and a BglII site in the extension of 3' primer. The PCR product was digested with BglII and cloned into BglII site of a human cytomeglovirus early promoter-based expression vector, VIJns (Figure 1A). The proper orientation of nef fragment in the context of the expression cassette was determined by asymmetric
 20 restriction mapping. The resultant plasmid is VIJns/nef. The 5' and 3' nucleotide sequence junctions of codon optimized VIJns/nef are shown in Figure 3A.

The mutant nef (G2A,LLAA) was also made from synthetic oligonucleotides. To assist in cloning, a PstI site and an SrfI site were included in the extensions of 5' and 3' PCR primers, respectively. The PCR product was digested with PstI and SrfI,
 25 and cloned into the PstI and SrfI sites of VIJns/nef, replacing the original nef with nef(G2A,LLAA) fragment. This resulted in VIJns/nef(G2A,LLAA). The 5' and 3' nucleotide sequence junctions of codon optimized VIJns/nef (G2A,LLAA) are shown in Figure 3B.

To construct the expression vector containing human tissue plasminogen
 30 activator leader peptide and the nef fusion gene, i.e., VIJns/tPAnef, a truncated nef gene fragment, lacking the coding sequence for the five amino terminal residues, was first amplified by PCR using VIJns/nef as template. Both 5' and 3' PCR primers used in this reaction contained a BglII extension. The PCR amplified fragment was then digested with BglII and cloned into BglII site of the expression vector, VIJns/tpa

(Figure 1B). The ligation of the 3' end of tpa leader peptide coding sequence to the 5' end of the nef PCR product restored the BglII site and yielded an in-frame fusion of the two genes. The 5' and 3' nucleotide sequence junctions of codon optimized V1Jns/tPAnef are shown in Figure 3C.

5 Construction of V1Jns/tpanef(LLAA) was carried out by replacing the Bsu36-SacII fragment of V1Jns/tpanef, which contains the 3' half of the nef gene and part of the vector backbone, with the Bsu36-SacII fragment from V1Jns/nef(G2A,LLAA). The 5' and 3' nucleotide sequence junctions of codon optimized V1Jns/tpanef (LLAA) are shown in Figure 3C.

10 All the nef constructs were verified by sequencing. The amino acid junctions of these constructs is shown schematically in Figure 4.

Transfection and protein expression - 293 cells (adenovirus transformed human embryonic kidney cell line 293) grown at approximately 30% confluence in minimum essential medium (MEM; GIBCO, Grand Island, MD) supplemented with
15 10% fetal bovine serum (FBS; GIBCO) in a 100 mm culture dish, were transfected with 4 ug gag expression vector, V1Jns/gag, or a mixture of 4 ug gag expression vector and 4 ug nef expression vector by Lipofectin following manufacture's protocol (GIBCO). Twelve hours post-transfection, cells were washed once with 10 ml of serum-free medium, Opti-MEM I (GIBCO) and replenished with 5 ml of Opti-MEM.
20 Following an additional 60 hr incubation, culture supernatants and cells were collected separately and used for Western blot analysis.

Western blot analysis - Fifty microliter of samples were separated on a 10% SDS-polyacrylamide gel (SDS-PAGE) under reducing conditions. The proteins were blotted onto a piece of PVDF membrane, and reacted to a mixture of gag mAb (#18; Intracel, Cambridge, MA) and Nef mAbs (aa64-68, aa195-201; Advanced
25 Biotechnologies, Columbia, MD), both at 1:2000 dilution, and horseradish peroxidase (HRP)-conjugated goat anti-rabbit IgG (Zymed, San Francisco, CA). The protein bands were visualized by ECL Western blotting detection reagents, according to the manufacture's protocol (Amersham, Arlington Heights, IL).

30 *Enzyme-linked immunosorbent assay (ELISA)* - 96-well Immulon II, round-bottom plates were coated with 50 ul of Nef protein at the concentration of 2ug/ml in bicarbonate buffer, pH 9.8., per well at 4°C overnight. Plates were washed three times with PBS containing 0.05% Tween-20 (PBST), and blocked with 5% skim milk in PBST (milk-PBST) at 24°C for 2 hr, and then incubated with serial dilutions of

testing samples in milk-PBST at 24°C for 2 hr. Plates were washed with PBST three times, and added with 50 ul of HRP-conjugated goat anti-mouse IgG (Zymed) per well and incubated at 24°C for 1 hr. This was followed by three washes, and the addition of 100 ul of 1 mg/ml ABTS [(2,2'-amino-di-(3-ethylbenzthiozoline sulfonate)] (KPL, Gaithersburg, MD) per well. After 1 hr at 24°C, plates were read at a wavelength of 405nm using an ELISA plate reader.

Enzyme-linked spot assay (Elispot) - Nitrocellulose membrane-backed 96 well plates (MSHA plates; Millipore, Bedford, MA) were coated with 50 ul of rat anti-mouse IFN-gamma mAb, capture antibody, (R4-6A2; PharMingen, San Diego, CA) at a concentration of 5ug/ml in PBS per well at 4°C overnight. Plates were washed three times with PBST and blocked with 10% FBS in RPMI-1640 (FBS-RPMI) at 37°C in a CO2 incubator for 2 to 4 hrs. Splenocytes were suspended in RPMI-1640 with 10% FBS at 4×10^6 cells per ml. 100 ul cells were added to each well and plates were incubated at 37°C for 20 hrs. Each sample was tested in triplicate wells. After incubation, plates were rinsed briefly with distilled water and washed three times with PBST. Fifty ul of biotinylated rat anti-mouse IFN- γ mAb, detecting antibody (XMG1.2; PharMingen), diluted in 1% BSA in PBST at a concentration of 2 ug/ml was then added to each well. Plates were incubated at 24°C for 2 hr, followed by washes with PBST. Fifty ul of streptavidin-conjugated alkaline phosphatase (KPL) at a dilution of 1:1000 in FBS-RPMI was added to each well. The plates were incubated at 24C for an additional one hr. Following extensive wash with BPST, 100ul BCIT/NBT substrate (KPL) was added for 15 min, and color reaction was stopped by washing the plate with tap water. Plates were air-dried and spots were countered using a dissection microscope.

Cytotoxic T cell (CTL) assay - Splenocytes from immunized mouse were co-cultured with syngenic peptide-pulsed, irradiated naive splenocytes for 7 days. EL-4 cells were incubated at 37°C for 1 hr with or without 20ug/ml of a designated peptide in the presence of sodium 51Cr-chromate and used as target cells. For the assay, 10^4 target cells were added to a 96-well plate along with different numbers of splenocytes cells. Plates were incubated at 37°C for 4 hr. After incubation, supernatants were collected and counted in a Wallac gamma-counter. Specific lysis was calculated as $[(\text{experimental release} - \text{spontaneous release}) / (\text{maximum release} - \text{spontaneous release})] \times 100\%$. Spontaneous release was determined by incubating target cells in

medium alone, and maximum release was determined by incubating target cells in 2.5% TritonX-100. The assay was performed with triplicate samples.

Animal experiments - Female mice (Charles River Laboratories, Wilmington, MA), 6 to 10 weeks old, were injected in quadriceps with 100 ul of DNA in PBS.

- 5 Two weeks after immunization, spleens from individual mice were collected and used for CTL and Elispot assays.

- Results (DNA Vector Vaccine Construction)* - The exemplified Nef protein sequence is based on HIV-1 clade B jrfl isolate. A codon-optimized nef gene was chosen for vaccine construction and for use as the parental gene for other exemplified constructs. Figure 2A-B show the comparison of coding sequence of wt nef(jrfl) and the codon optimized nef(jrfl). Two forms of myristylation site mutations were constructed; one contains a Gly2Ala change and the other a human tissue plasminogen activator (tpa) leader sequence was fused to sixth residue, Ser, of Nef (tpanef). The dileucine motif mutation was made by introducing both Leu174Ala and
10 Leu175Ala changes. Figure 4 shows the schematic depiction of the Nef and Nef mutants. For *in vitro* expression and *in vivo* immunogenicity studies, the nef genes were cloned into expression vector, V1Jns. The resultant plasmids containing wt nef, tpanef, tpanef with dileucine motif mutation, and nef mutant with the Gly2Ala myristylation site and dileucine motif mutations were named as V1Jns/nef,
15 V1Jns/tpanef, V1Jns/tpanef(LLAA) and V1Jns/(G2A,LLAA), respectively.

- Results - Expression and Western blotting analysis* - To evaluate the expression of the codon optimized nef constructs, adenovirus-transformed human kidney 293 cells were cotransfected with individual nef plasmids and a gag expression vector, V1Jns/gag. 72 hours post transfection, cells and medium were collected
20 separately and analyzed by Western blotting, using both Nef- and Gag-specific mAbs. The results are shown in Figure 5. Cells transfected with V1Jns/gag only revealed a single distinct band of approximately 55 Kd, whereas the cells cotransfected with gag and nef plasmids revealed, in addition to the 55 Kd band, a major 30 Kd band and several minor bands. This pattern is consistent with that the 55 Kd species represents
25 Gag polypeptide and the 30 Kd and other minor species are the Nef-related products. Therefore, all the nef constructs were expressed in the transfected cells. When measured against the relatively constant Gag signal as a reference, four nef genes seem to be expressed at different levels, with the following descending order, tpanef, nef, tpanef(LLAA) and nef(G2A,LLAA). With the exception of nef(G2A,LLAA),
30

products of nef, tpanef, tpanef(LLAA) could be detected in both cellular and medium fractions.

Mapping of Nef-specific CD8 and CD4 epitopes in mice – There was no information available with respect to the properties of Nef(jrfl) in eliciting cell-mediated immune responses in mice. Therefore, to characterize immunogenicity of Nef and Nef mutants exemplified herein, CD8 and CD4 epitopes were mapped. An overlapping set of overlapping nef peptides that encompass the entire 216 aa Nef polypeptide were generated. A total 21 peptides were made, which include twenty 20mers and one 16mer. Three strains of mice, Balb/c, C3H and C57BL/6, were immunized with plasmid V1Jns/Nef; splenocytes from immunized and naive mice were isolated and assessed for Nef specific INF-gamma secreting cells (SFC) by the Elispot assay. Figure 6 shows where Elispot assays were performed against separate pools of the Nef peptides. All three strains of immunized mice responded to the Nef plasmid immunization; each developed positive Nef peptide-specific INF- γ SFCs. Based on this, further studies were carried out with fractionated CD8 and CD4 cells against individual peptides. The results are shown in Figure 7A-C. In Balb/c mice (Figure 7A), four Nef peptides, namely, aa11-30, aa61-80, aa191-210 and aa200-216, were found to be able to induce significant numbers of CD4 SFCs. In C57BL/6 mice (Figure 7B), only one peptide, ie., aa81-100, elicited significant numbers of CD4 SFCs. Compared to Balb/c and C57BL/6 mice, C3H mice (Figure 7C) showed no dominant CD4 SFC responses with particular peptides; instead, there were modest number of SFCs in response to an array of peptides, including aa21-40, aa31-50, aa121-140 aa131-150, aa181-200 and aa191-210. With respect to CD8 cells, significant SFC responses were detected with a single peptide, ie., aa51-70, in C57BL/6 mice only.

The results from Elispot assay suggested that Nef peptide aa51-70 contained an H-2b restricted CD8 cell epitope. In order to ascertain whether this CD8 epitope also represents the cytotoxic T cell (CTL) epitope, a conventional CTL assay was carried out. The peptide aa51-70 (Figure 8A) induced low level of specific killings only. Peptides longer than 9 amino acids of a typical CTL epitope often have lower binding affinity to MHC class I molecule. It was contemplated that the low specific killings observed with peptide aa51-70 could be potentially resulted from the low binding affinity of this 20 amino acid peptide. Therefore, two shortened peptides, namely, aa60-68 and aa58-70, were synthesized and tested in CTL assays. While the

peptide aa60-68 failed to elicit any specific killings (Figure 8B), the peptide aa58-70 exhibited a drastic increase of specific killing as compared to its longer counterpart, peptide aa61-80 (Figure 8C). For example, the percentage of specific killings induced by peptide aa58-70 at an effector/target ratio of 5 to 1 was comparable to that induced by peptide aa51-80 at an effector/target ratio of 45. Thus, between peptide aa58-70 and peptide aa51-70, the former was almost ten-fold more effective in terms of inducing Nef-specific killing. The results from CTL assay therefore confirmed that the CD8 epitope detected by the Elispot assay was indeed a CTL epitope. To further map the minimum amino acid sequence for the Nef CTL epitope, additional 5 peptides were synthesized and analyzed by Elispot assay, which mapped the CTL epitope to Nef aa58-66, as shown in Table 1.

TABLE 1

Nef peptides**		INF- γ SFC*/10 ⁶ splenocytes
Nef58-70	TAATNADCAWLEA	85
Nef59-69	AATNADCAWLE	1
Nef58-68	TAATNADCAWL	69
Nef58-67	TAATNADCAW	66
Nef58-66	TAATNADCA	92
Medium		1

* Average of duplicate samples.

15 ** Amino acid sequence of all peptides contained within SEQ ID NO:2.

Results (Evaluation of Immunogenicity of nef Mutants in Mice) - Having identified H-2b restricted CTL and CD4 cell epitopes, the immunogenicity of the different codon optimized nef constructs in C57BL/6 mice was examined. This was performed in two separate experiments with identical immunization regimens. The first experiment involved nef, tpanef(LLAA) and nef(G2A,LLAA) and the second experiment involved nef, tpanef, tpanef(LLAA) and nef(G2A,LLAA). Mice were immunized with plasmids containing these respective codon optimized nef genes. Two weeks post immunization, splenocytes from individual mice were isolated and analyzed by Elispot assay for Nef-specific CD8 and CD4 IFN-gamma SFCs using Nef peptide aa58-66 and aa81-100, respectively. The results are shown in Figure 9A-B. In the experiment 1 (Figure 9A), among the three groups tested, the mice receiving the codon optimized tpanef(LLAA) construct developed the highest CD8 and CD4 cell responses; comparing between tpanef(LLAA) and the nef, the former elicited about 40-fold higher CD8 SFCs and 10-fold higher CD4 SFCs. In contrast to tpanef(LLAA), nef(G2A,LLAA) mutant was poorly immunogenic; mice receiving this mutant had barely detectable CD8 and CD4 SFCS, under conditions tested. Similar response profiles between the three mutants were also observed in the experiment 2 (Figure 9B), except that the overall CD8 response of mice receiving tpanef(LLAA) was approximately 10-fold higher in experiment 2 than that observed in experiment 1. The tPAnef mutant showed comparable responses as that of tpanef(LLAA). The results therefore showed that both codon optimized tpanef and tpanef(LLAA) had significantly enhanced immunogenicity.

Results (Evaluation of Immunogenicity of nef Mutants in Rhesus Monkeys) - Monkeys were immunized with 5 mg of indicated codon optimized plasmids at week 0, 4, and 8. Four weeks after each immunization, peripheral blood mononuclear cells were collected and tested for Nef-specific INF-gamma secreting cells as described for the mice studies in this Example section. The results are shown in Table 2. As with the mouse study, tpanef(LLAA) shows significantly enhanced immunogenicity when compared to tPAnef.

TABLE 2

Vaccine	Animal No.	Nef specific INF-gamma secreting cells/million PBMC							
		Week 0		Week 4		Week 8		Week 12	
		Medium	nef	Medium	nef	Medium	nef	Medium	nef
V1Jns-TpaNef (LLAA)	1	74	39	30	208	6	148	89	559
	2	1	3	28	45	13	44	13	146
	3	5	5	14	45	11	11	14	35
V1Jns-nef	1	0	1	24	33	16	43	6	34
	2	28	9	31	35	13	34	24	80
	3	1	0	16	31	18	38	13	185
Control	1	1	3	16	33	16	16	18	13

- Monkeys were immunized with 5 mg of indicated plasmids at week 0, 4 and 8.
- 5 Four weeks after each immunization, peripheral blood mononuclear cells were collected and tested for the Nef-specific IFN-gamma secreting cells.

A codon-optimized nef gene coding for HIV-1 jrf1 isolate Nef polypeptide was synthesized. The resultant synthetic nef gene was well expressed in the *in vitro* transfected cells. Using this synthetic gene as parental molecule, nef mutants involving myristylation site and dileucine motif mutations were constructed. Two forms of myristylation site mutation were made, one involving a single Gly2Ala change and the other by fusing human plasminogen activator(tpa) leader peptide with the N-terminus of Nef polypeptide. The dileucine motif mutation was generated by Leu174Ala and Leu175Ala changes. The resultant nef constructs were named as nef, tpanef, tpanef(LLAA) and nef(G2A,LLAA). The addition of tpa leader peptide sequence resulted in significantly increased expression of the nef gene *in vitro*; in contrast, either Gly2Ala mutation or dileucine mutation reduced the nef gene

expression. In an effort to characterize immunogenicity of nef and nef mutants, experiments were carried out to map nef CTL and Th epitopes in mice. A single CTL epitope and a dominant Th epitope, both restricted by H-2b, were identified. Consequently, C57BL/6 mice were immunized with different nef constructs by DNA
5 immunization means, and splenocytes from immunized mice were determined for Nef-specific CTL and Th responses using Elisopt assay and the defined T cell epitopes. The results showed that tpanef and tpanef(LLAA) were significantly more immunogenic than nef in terms of eliciting both CTL and Th responses.

Therefore, these aforementioned polynucleotides, when directly introduced
10 into a vertebrate *in vivo*, including mammals such as primates and humans, should express the respective HIV-1 Nef protein within the animal and in turn induce at least a cytotoxic T lymphocyte (CTL) response within the host to the expressed Nef antigen.

The present invention is not to be limited in scope by the specific
15 embodiments described herein. Indeed, various modifications of the invention in addition to those described herein will become apparent to those skilled in the art from the foregoing description. Such modifications are intended to fall within the scope of the appended claims.

20

WHAT IS CLAIMED IS:

1. A pharmaceutically acceptable DNA vaccine, which comprises:
 - (a) a DNA expression vector; and,
 - 5 (b) a DNA molecule containing a codon optimized open reading frame encoding a Nef protein or immunogenic Nef derivative thereof, wherein upon administration of the DNA vaccine to a host the Nef protein or immunogenic Nef derivative is expressed and generates an immune response which provides a substantial level of protection against HIV-1 infection.
- 10 2. A DNA vaccine of claim 1 wherein the DNA molecule encodes wild type Nef.
3. A DNA vaccine of claim 2 wherein the DNA molecule contains the
15 nucleotide sequence as set forth in SEQ ID NO:1.
4. The DNA vaccine of claim 3 which is V1Jns-opt nef (jrfl).
5. A DNA vaccine of claim 2 wherein the DNA molecule expresses a
20 wild type Nef protein which comprises the amino acid sequence as set forth in SEQ ID NO:2.
6. A DNA vaccine of claim 1 wherein the DNA molecule encodes an
immunogenic Nef derivative which contains a nucleotide sequence encoding a leader
25 peptide.
7. A DNA vaccine of claim 6 wherein the DNA molecule encodes an
immunogenic Nef derivative which contains a nucleotide sequence encoding a human
tissue plasminogen activator leader peptide.
30
8. A DNA vaccine of claim 7 wherein the DNA molecule contains the
nucleotide sequence as set forth in SEQ ID NO:3.
9. The DNA vaccine of claim 8 which is V1Jns-opt tpanef.

10. A DNA vaccine of claim 7 wherein the DNA molecule expresses an immunogenic Nef derivative which comprises the amino acid sequence as set forth in SEQ ID NO:4.

5

11. A DNA vaccine of claim 6 wherein the DNA molecule encodes an immunogenic Nef derivative modified at the dileucine motif of amino acid residue 174 and amino acid residue 175.

10

12. A DNA vaccine of claim 11 wherein the DNA molecule encodes an immunogenic Nef derivative which contains a nucleotide sequence encoding a human tissue plasminogen activator leader peptide.

13. A DNA vaccine of claim 12 wherein the DNA molecule contains the nucleotide sequence as set forth in SEQ ID NO:7.

15

14. The DNA vaccine of claim 13 which is V1Jns-opt tpanef (LLAA).

15. A DNA vaccine of claim 11 wherein the DNA molecule expresses an immunogenic Nef derivative which comprises the amino acid sequence as set forth in SEQ ID NO:8.

20

16. A DNA vaccine of claim 11 wherein the DNA molecule encodes a Nef protein where the glycine residue of amino acid residue 2 of Nef is modified to encode for an amino acid residue other the glycine.

25

17. A DNA vaccine of claim 16 wherein the DNA molecule contains the nucleotide sequence as set forth in SEQ ID NO:5.

30

18. A DNA vaccine of claim 17 which is V1Jns-opt nef (G2A LLAA).

19. A DNA vaccine of claim 16 wherein the DNA molecule expresses an immunogenic Nef derivative which comprises the amino acid sequence as set forth in SEQ ID NO:6.

20. A DNA vaccine of claim 1 which further comprises an adjuvant.
21. A DNA vaccine of claim 20 wherein the adjuvant is selected from the
5 group consisting of aluminum phosphate, calcium phosphate and a non-ionic block copolymer.
22. A pharmaceutically acceptable DNA vaccine, which comprises:
(a) a DNA expression vector; and,
10 (b) a DNA molecule containing an open reading frame encoding a Nef protein or immunogenic Nef derivative thereof, wherein upon administration of the DNA vaccine to a host the Nef protein or immunogenic Nef derivative is expressed and generates an immune response which provides a substantial level of protection against HIV-1 infection.
15
23. The DNA vaccine of claim 22 wherein the DNA molecule expresses a wild type Nef protein which comprises the amino acid sequence as set forth in the group consisting of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6 and SEQ ID NO:8.
- 20 24. A DNA vaccine of claim 22 which further comprises an adjuvant.
25. A DNA vaccine of claim 23 wherein the adjuvant is selected from the group consisting of aluminum phosphate, calcium phosphate and a non-ionic block copolymer.
25
26. A method for inducing a cell mediated immune (CTL) response against infection or disease caused by virulent strains of HIV which comprises administering into the tissue of a vertebrate host a pharmaceutically acceptable DNA vaccine composition which comprises a DNA expression vector and a DNA molecule
30 containing a codon optimized open reading frame encoding a Nef protein or immunogenic Nef derivative thereof, wherein upon administration of the DNA vaccine to the vertebrate host the Nef protein or immunogenic Nef derivative is expressed and generates the cell-mediated immune (CTL) response.

27. The method of claim 26 wherein the vertebrate host is a human.

28. The method of claim 26 wherein the DNA vaccine is selected from the group consisting of V1Jns-opt nef (jrf1), V1Jns-opt tpanef, V1Jns-opt tpanef (LLAA),
5 and V1Jns-opt nef (G2A LLAA).

29. A substantially purified protein which comprises an amino acid sequence selected from the group consisting of SEQ ID NO:4, SEQ ID NO:6, and SEQ ID NO:8.

10

1/10

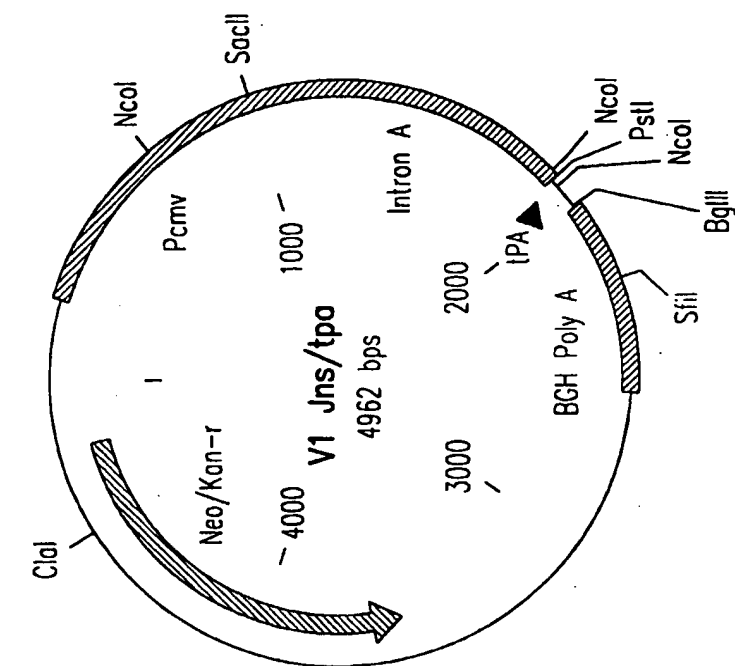


FIG.1B

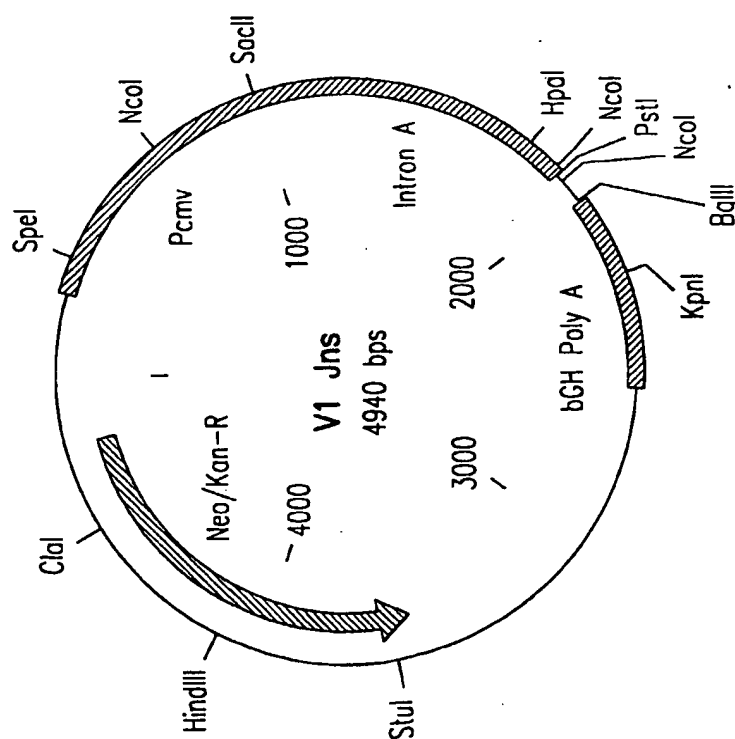


FIG.1A

2/10

WT	- ATG GGT GGC AAG TGG TCA AAA CGT AGT GTG CCT GGA TGG TCT	-42
OPT	- ATG GGC GGC AAG TGG TCC AAG AGG TCC GTG CCC GGC TGG TCC	
	M G G K W S K R S V P G W S	-14
WT	- ACT GTA AGG GAA AGA ATG AGA CGA GCT GAG CCA GCA GCA GAT	-84
OPT	- ACC GTG AGG GAG AGG ATG AGG AGG GCC GAG CCC GCC GCC GAC	
	T V R E R M R R A E P A A D	-28
WT	- AGG GTG AGA CGA ACT GAG CCA GCA GCA GTA GGG GTG GGA GCA	-126
OPT	- AGG GTG AGG AGG ACC GAG CCC GCC GCC GTG GGC GTG GGC GCC	
	R V R R T E P A A V G V G A	-42
WT	- GTA TCT CGA GAC CTG GAA AAA CAT GGA GCA ATC ACA AGT AGC	-168
OPT	- GTG TCC AGG GAC CTG GAG AAG CAC GGC GCC ATC ACC TCC TCC	
	V S R D L E K H G A I T S S	-56
WT	- AAT ACA GCA GCT ACC AAT GCT GAT TGT GCC TGG CTA GAA GCA	-210
OPT	- AAC ACC GCC GCC ACC AAC GCC GAC TGC GCC TGG CTG GAG GCC	
	N T A A T N A D C A W L E A	-70
WT	- CAA GAG GAT GAG GAA GTG GGT TTT CCA GTC AGA CCT CAG GTA	-252
OPT	- CAG GAG GAC GAG GAG GTG GGC TTC CCC GTG AGG CCC CAG GTG	
	Q E D E E V G F P V R P Q V	-84
WT	- CCT TTA AGA CCA ATG ACT TAC AAG GGA GCT GTA GAT CTT AGC	-294
OPT	- CCC CTG AGG CCC ATG ACC TAC AAG GGC GCC GTG GAC CTG TCC	
	P L R P M T Y K G A V D L S	-98
WT	- CAC TTT TTA AAA GAA AAG GGG GGA CTG GAA GGG CTA ATT CAC	-336
OPT	- CAC TTC CTG AAG GAG AAG GGC GGC CTG GAG GGC CTG ATC CAC	
	H F L K E K G G L E G L I H	-112
WT	- TCA CAG AAA AGA CAA GAT ATC CTT GAT CTG TGG GTC TAC CAC	-378
OPT	- TCC CAG AAG AGG CAG GAC ATC CTG GAC CTG TGG GTG TAC CAC	
	S Q K R Q D I L D L W V Y H	-126
WT	- ACA CAA GGC TAC TTC CCT GAT TGG CAG AAC TAC ACA CCA GGG	-420
OPT	- ACC CAG GGC TAC TTC CCC GAC TGG CAG AAC TAC ACC CCC GGC	
	T Q G Y F P D W Q N Y T P G	-140

FIG.2A

3/10

WT	- CCA GGA ATC AGA TTT CCA TTG ACC TTT GGA TGG TGC TTC AAG	-462
OPT	- CCC GGC ATC AGG TTC CCC CTG ACC TTC GGC TGG TGC TTC AAG	
	P G I R F P L T F G W C F K	-154
WT	- CTA GTA CCA GTT GAG CCA GAA AAG GTA GAA GAG GCC AAT GAA	-504
OPT	- CTG GTG CCC GTG GAG CCC GAG AAG GTG GAG GAG GCC AAC GAG	
	L V P V E P E K V E E A N E	-168
WT	- GGA GAG AAC AAC TGC TTG TTA CAC CCT ATG AGC CAG CAT GGG	-546
OPT	- GGC GAG AAC AAC TGC CTG CTG CAC CCC ATG TCC CAG CAC GGC	
	G E N N C L L H P M S Q H G	-182
WT	- ATA GAG GAC CCG GAG AAG GAA GTG TTA GAG TGG AGG TTT GAC	-588
OPT	- ATC GAG GAC CCC GAG AAG GAG GTG CTG GAG TGG AGG TTC GAC	
	I E D P E K E V L E W R F D	-196
WT	- AGC AAG CTA GCA TTT CAT CAC GTG GCC CGA GAG CTG CAT CCG	-630
OPT	- TCC AAG CTG GCC TTC CAC CAC GTG GCC AGG GAG CTG CAC CCC	
	S K L A F H H V A R E L H P	-210
WT	- GAG TAC TAC AAG GAC TGC TGA (SEQ ID NO:30)	-651
OPT	- GAG TAC TAC AAG GAC TGC TAA (contained within SEQ ID NO:1)	
	E Y Y K D C (SEQ ID NO:2)	-216

FIG.2B

4/10

VIJns/nef *PstI* *BglIII*
CATGGGCTCTTTTCTGCAGTCACCGTCCTTGGAGATCTGCCACC ATG GGC GGC AAG TGG TCC AAG AGG TCC GTG CCC
M G G K W S K R S V P
. CAC CCC GAG TAC TAC AAG GAC TGC TAA AGCCCCGGGCAGATCTGCTGTGCTTCTAGTTGCCAGC (SEQ ID NO:27)
H P E Y Y K D C * (contained within SEQ ID NO:2)

FIG.3A

VIJns/nef(G2A,LLAA)
PstI *BglIII*
CATGGGCTCTTTTCTGCAGTCACCGTCCTTGGAGATCTGCCACC ATG GCC GGC AAG TGG TCC AAG AGG TCC GTG CCC
M A G K W S K R S V P
. CAC CCC GAG TAC TAC AAG GAC TGC TAA AGCCCCGGGCAGATCTGCTGTGCTTCTAGTTGCCAGC (SEQ ID NO:28)
H P E Y Y K D C * (contained within SEQ ID NO:6)

FIG.3B

VIJns/tpanef & VIJns/tpanef(LLAA)
PstI *BglIII*
CATGGGCTCTTTTCTGCAGTCACCGTCCTTATATCTAGATCACC ATG GAT GCA ATG AAG AGA GGG CTC TGC TGT GTG
M D A M K R G L C C V
CTG CTG CTG TGT GGA GCA GTC TTC GTT TCG CCC AGC GAGATC ICC TCC AAG AGG TCC GTG CCC
L L L C G A V F V S P S E I S S K R S V P
. CAC CCC GAG TAC TAC AAG GAC TGC TAA AGCCCCGGGCAGATCTGCTGTGCTTCTAGTTGCCAGC (SEQ ID NO:29)
H P E Y Y K D C * (contained within SEQ ID NO:8)

FIG.3C

5/10

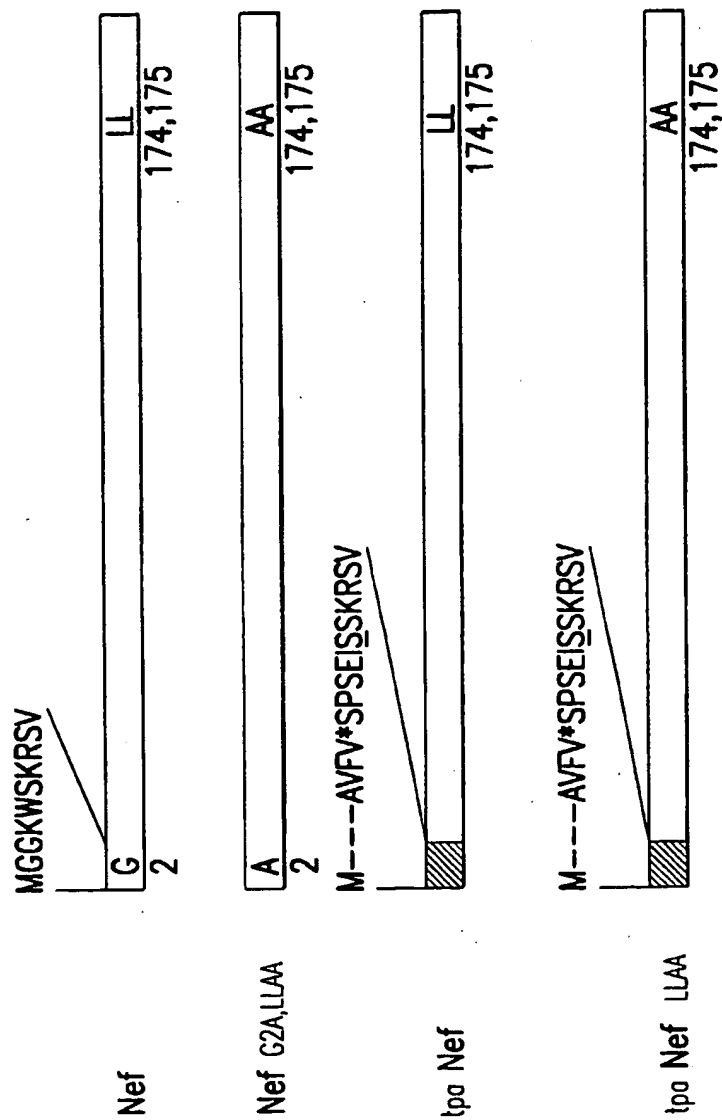


FIG.4

6/10

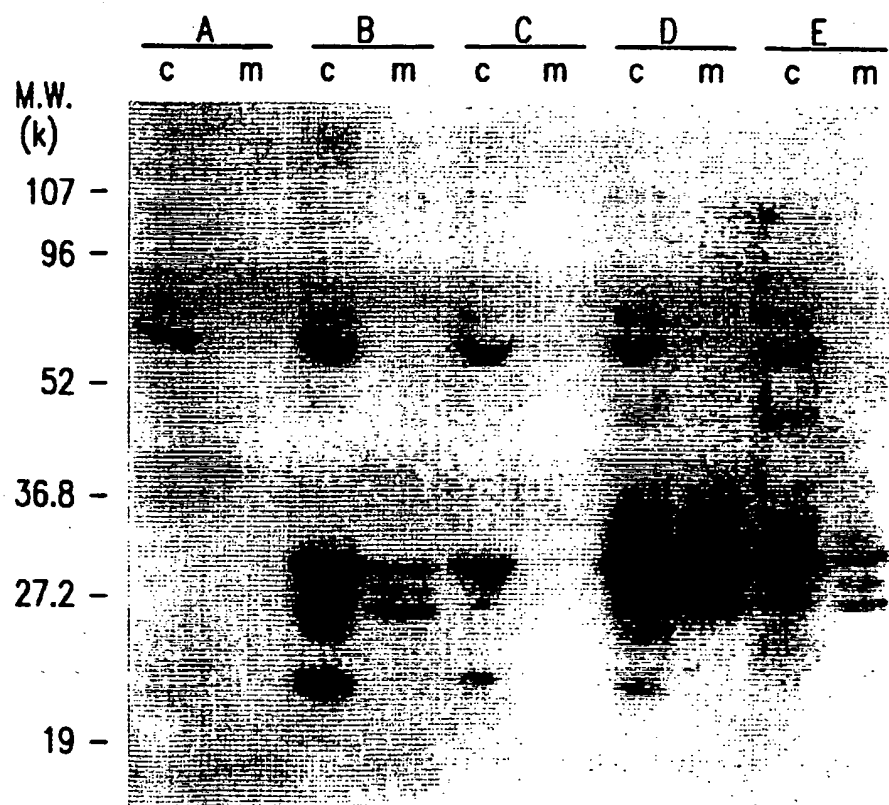


FIG.5

7/10

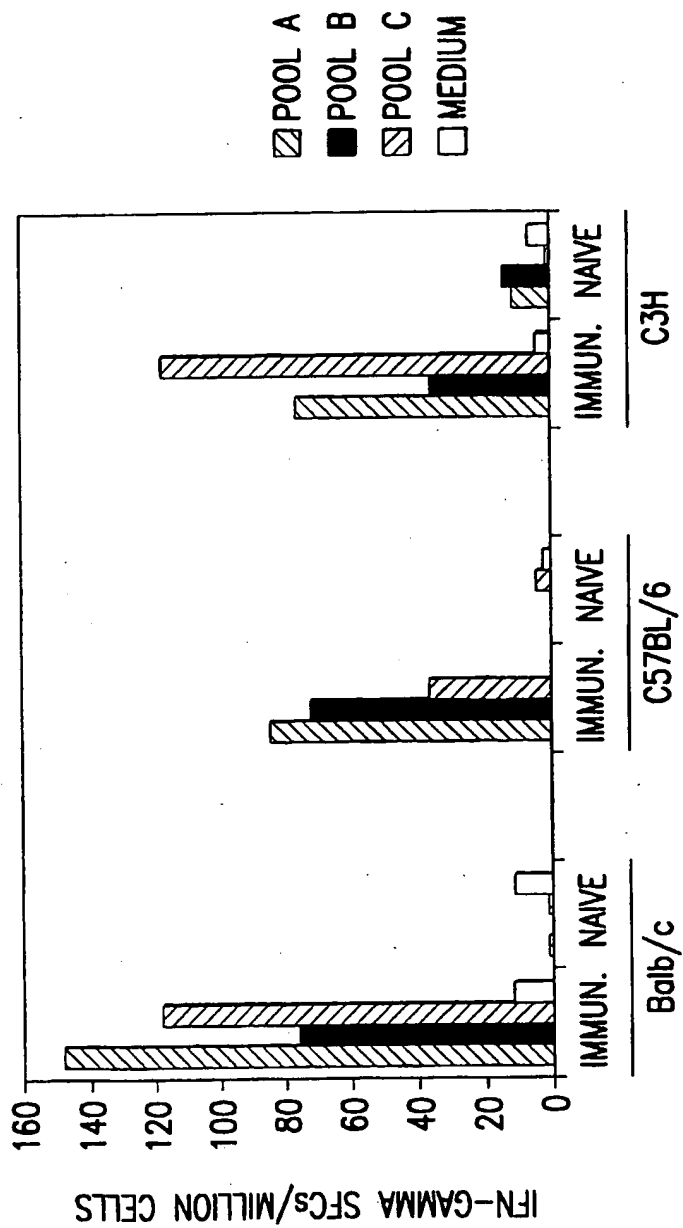
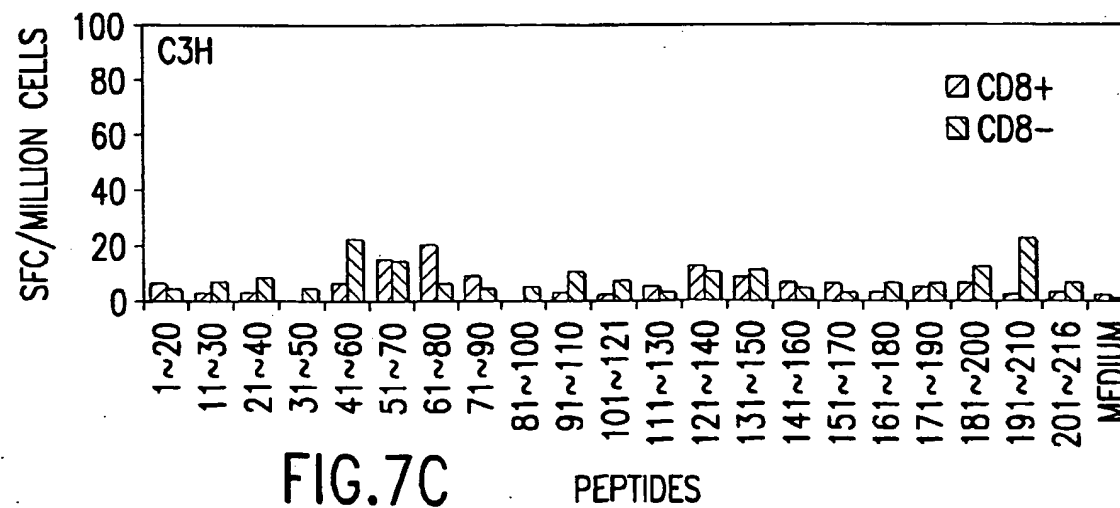
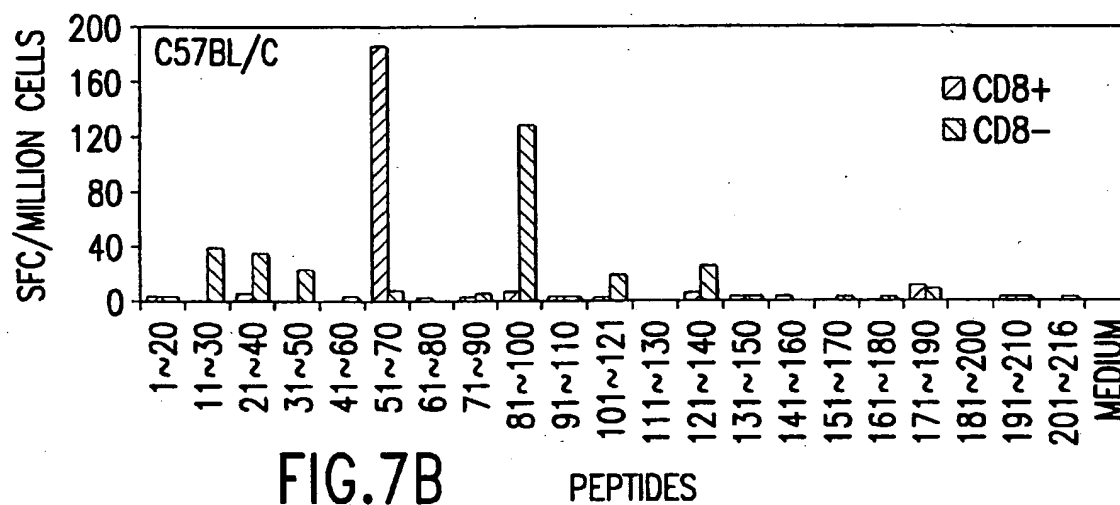
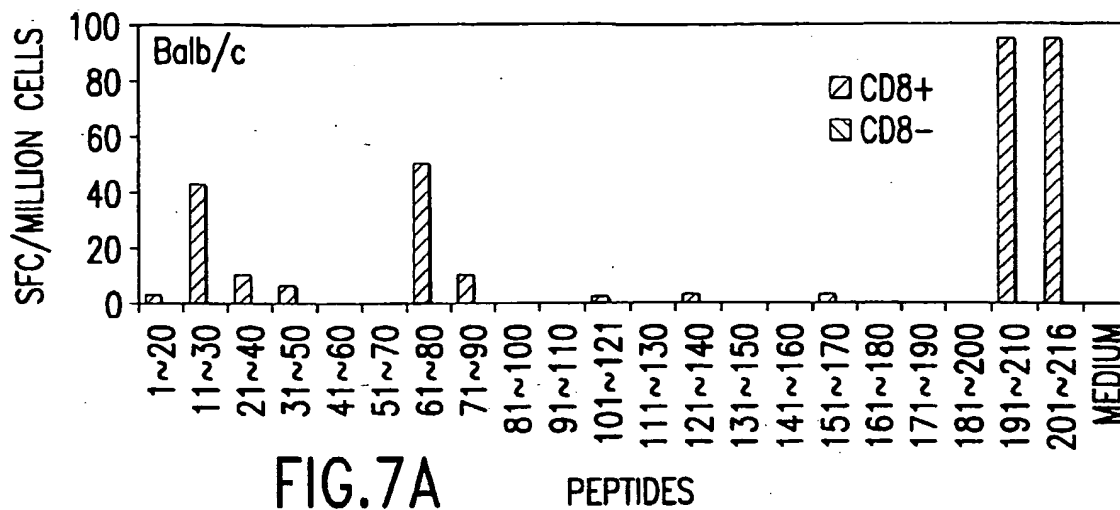


FIG.6

8/10



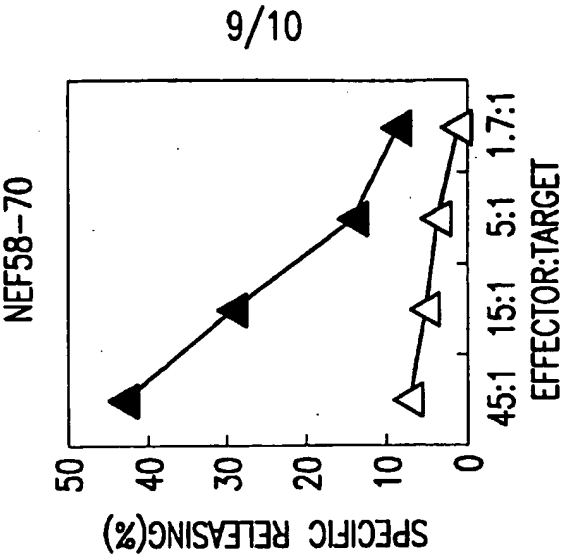


FIG.8C

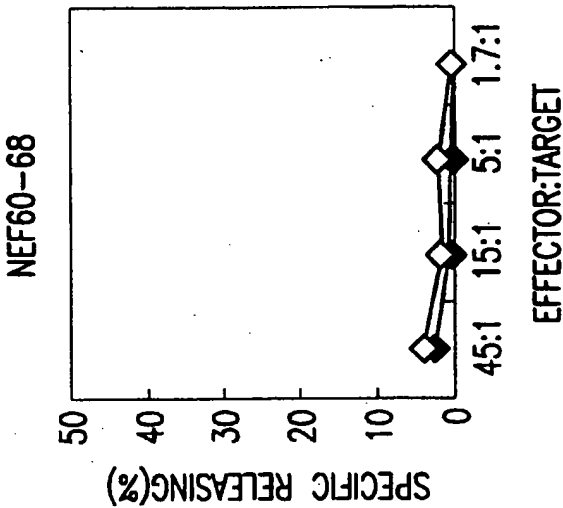


FIG.8B

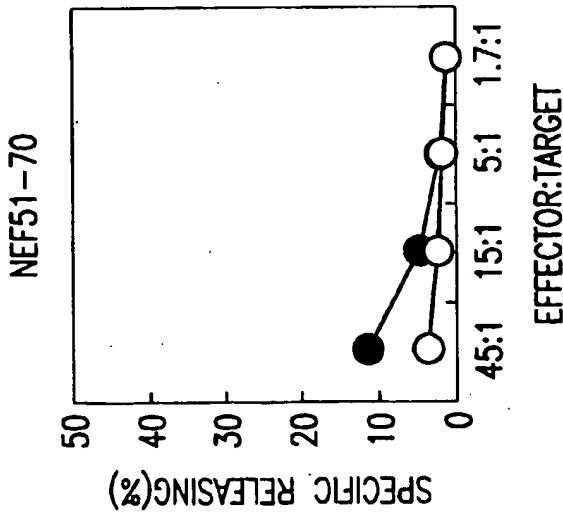


FIG.8A

10/10

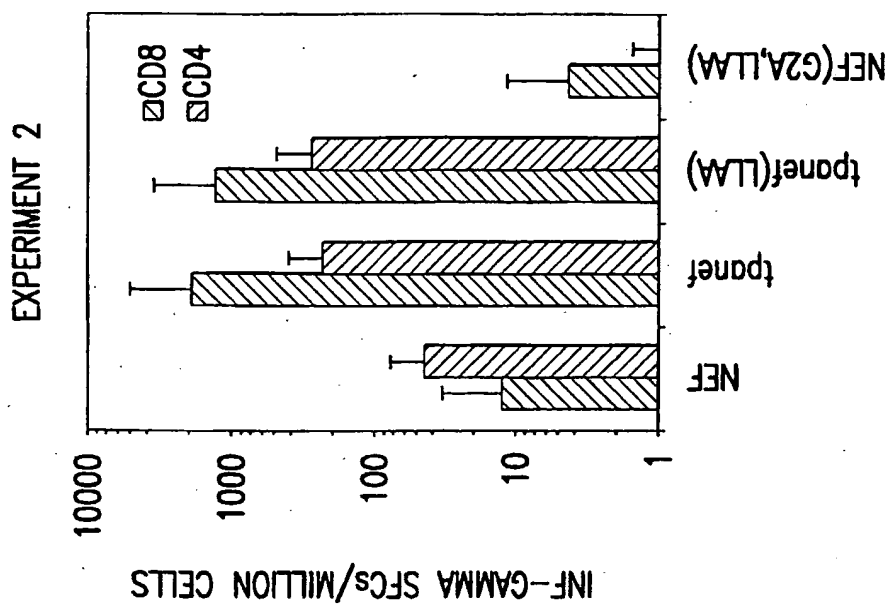


FIG.9B

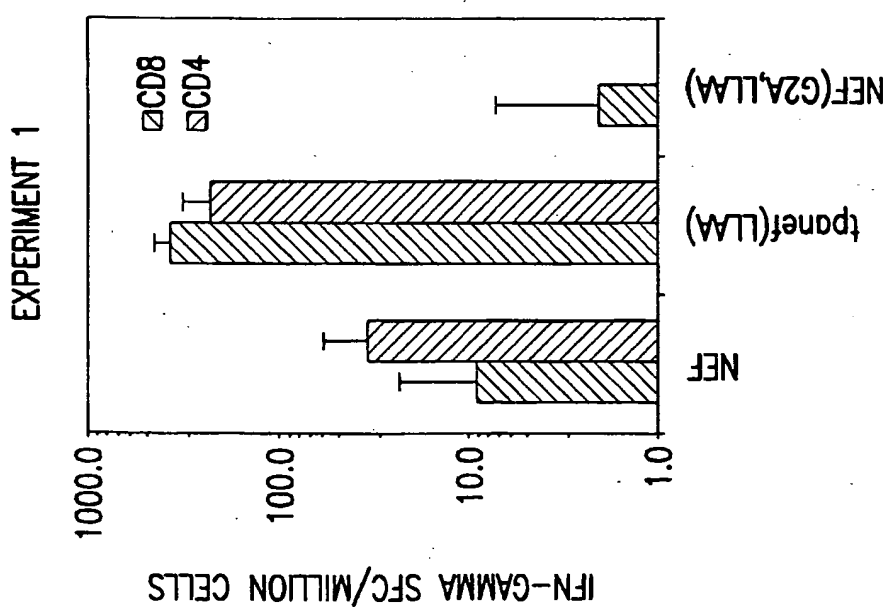


FIG.9A

SEQUENCE LISTING

<110> APPLICANT: Merck & Co., Inc.

<120> TITLE: POLYNUCLEOTIDE VACCINES EXPRESSING CODON
OPTIMIZED HIV-1 NEF AND MODIFIED HIV-1 NEF

<130> DOCKET/FILE REFERENCE: 20602Y

<160> NUMBER OF SEQUENCES: 30

<170> SOFTWARE: FastSEQ for Windows Version 4.0

<210> SEQ ID NO:1

<211> LENGTH: 671

<212> TYPE: DNA

<213> ORGANISM: Human Immunodeficiency Virus - 1

<220> FEATURE:

<221> NAME/KEY: CDS

<222> LOCATION: (12)...(662)

<400> SEQ ID NO:1

```

gatctgccac c atg ggc ggc aag tgg tcc aag agg tcc gtg ccc ggc tgg      50
      Met Gly Gly Lys Trp Ser Lys Arg Ser Val Pro Gly Trp
          1              5              10

tcc acc gtg agg gag agg atg agg agg gcc gag ccc gcc gcc gac agg      98
Ser Thr Val Arg Glu Arg Met Arg Arg Ala Glu Pro Ala Ala Asp Arg
      15              20              25

gtg agg agg acc gag ccc gcc gcc gtg ggc gtg ggc gcc gtg tcc agg      146
Val Arg Arg Thr Glu Pro Ala Ala Val Gly Val Gly Ala Val Ser Arg
      30              35              40              45

gac ctg gag aag cac ggc gcc atc acc tcc tcc aac acc gcc gcc acc      194
Asp Leu Glu Lys His Gly Ala Ile Thr Ser Ser Asn Thr Ala Ala Thr
          50              55              60

aac gcc gac tgc gcc tgg ctg gag gcc cag gag gac gag gag gtg ggc      242
Asn Ala Asp Cys Ala Trp Leu Glu Ala Gln Glu Asp Glu Glu Val Gly
          65              70              75

ttc ccc gtg agg ccc cag gtg ccc ctg agg ccc atg acc tac aag ggc      290
Phe Pro Val Arg Pro Gln Val Pro Leu Arg Pro Met Thr Tyr Lys Gly
          80              85              90

gcc gtg gac ctg tcc cac ttc ctg aag gag aag ggc ggc ctg gag ggc      338
Ala Val Asp Leu Ser His Phe Leu Lys Glu Lys Gly Gly Leu Glu Gly
          95              100              105

ctg atc cac tcc cag aag agg cag gac atc ctg gac ctg tgg gtg tac      386
Leu Ile His Ser Gln Lys Arg Gln Asp Ile Leu Asp Leu Trp Val Tyr
      110              115              120              125

cac acc cag ggc tac ttc ccc gac tgg cag aac tac acc ccc ggc ccc      434
His Thr Gln Gly Tyr Phe Pro Asp Trp Gln Asn Tyr Thr Pro Gly Pro

```

130	135	140	
ggc atc agg ttc ccc ctg acc ttc ggc tgg tgc ttc aag ctg gtg ccc			482
Gly Ile Arg Phe Pro Leu Thr Phe Gly Trp Cys Phe Lys Leu Val Pro			
145	150	155	
gtg gag ccc gag aag gtg gag gag gcc aac gag ggc gag aac aac tgc			530
Val Glu Pro Glu Lys Val Glu Glu Ala Asn Glu Gly Glu Asn Asn Cys			
160	165	170	
ctg ctg cac ccc atg tcc cag cac ggc atc gag gac ccc gag aag gag			578
Leu Leu His Pro Met Ser Gln His Gly Ile Glu Asp Pro Glu Lys Glu			
175	180	185	
gtg ctg gag tgg agg ttc gac tcc aag ctg gcc ttc cac cac gtg gcc			626
Val Leu Glu Trp Arg Phe Asp Ser Lys Leu Ala Phe His His Val Ala			
190	195	200	205
agg gag ctg cac ccc gag tac tac aag gac tgc taa agccccgggc			671
Arg Glu Leu His Pro Glu Tyr Tyr Lys Asp Cys *			
210	215		

<210> SEQ ID NO:2

<211> LENGTH: 216

<212> TYPE: PRT

<213> ORGANISM:Human Immunodeficiency Virus - 1

<400> SEQ ID NO:2

Met Gly Gly Lys Trp Ser Lys Arg Ser Val Pro Gly Trp Ser Thr Val			
1 5 10 15			
Arg Glu Arg Met Arg Arg Ala Glu Pro Ala Ala Asp Arg Val Arg Arg			
20 25 30			
Thr Glu Pro Ala Ala Val Gly Val Gly Ala Val Ser Arg Asp Leu Glu			
35 40 45			
Lys His Gly Ala Ile Thr Ser Ser Asn Thr Ala Ala Thr Asn Ala Asp			
50 55 60			
Cys Ala Trp Leu Glu Ala Gln Glu Asp Glu Glu Val Gly Phe Pro Val			
65 70 75 80			
Arg Pro Gln Val Pro Leu Arg Pro Met Thr Tyr Lys Gly Ala Val Asp			
85 90 95			
Leu Ser His Phe Leu Lys Glu Lys Gly Gly Leu Glu Gly Leu Ile His			
100 105 110			
Ser Gln Lys Arg Gln Asp Ile Leu Asp Leu Trp Val Tyr His Thr Gln			
115 120 125			
Gly Tyr Phe Pro Asp Trp Gln Asn Tyr Thr Pro Gly Pro Gly Ile Arg			
130 135 140			
Phe Pro Leu Thr Phe Gly Trp Cys Phe Lys Leu Val Pro Val Glu Pro			
145 150 155 160			
Glu Lys Val Glu Glu Ala Asn Glu Gly Glu Asn Asn Cys Leu Leu His			
165 170 175			
Pro Met Ser Gln His Gly Ile Glu Asp Pro Glu Lys Glu Val Leu Glu			
180 185 190			
Trp Arg Phe Asp Ser Lys Leu Ala Phe His His Val Ala Arg Glu Leu			
195 200 205			
His Pro Glu Tyr Tyr Lys Asp Cys			
210 215			

<210> SEQ ID NO:3

<211> LENGTH: 719

<212> TYPE: DNA

<213> ORGANISM: Human Immunodeficiency Virus - 1

<220> FEATURE:

<221> NAME/KEY: CDS

<222> LOCATION: (2)...(715)

<400> SEQ ID NO:3

```

c atg gat gca atg aag aga ggg ctc tgc tgt gtg ctg ctg ctg tgt gga      49
Met Asp Ala Met Lys Arg Gly Leu Cys Cys Val Leu Leu Leu Cys Gly
  1             5             10             15

gca gtc ttc gtt tcg ccc agc gag atc tcc tcc aag agg tcc gtg ccc      97
Ala Val Phe Val Ser Pro Ser Glu Ile Ser Ser Lys Arg Ser Val Pro
      20             25             30

ggc tgg tcc acc gtg agg gag agg atg agg agg gcc gag ccc gcc gcc      145
Gly Trp Ser Thr Val Arg Glu Arg Met Arg Arg Ala Glu Pro Ala Ala
      35             40             45

gac agg gtg agg agg acc gag ccc gcc gcc gtg ggc gtg ggc gcc gtg      193
Asp Arg Val Arg Arg Thr Glu Pro Ala Ala Val Gly Val Gly Ala Val
      50             55             60

tcc agg gac ctg gag aag cac ggc gcc atc acc tcc tcc aac acc gcc      241
Ser Arg Asp Leu Glu Lys His Gly Ala Ile Thr Ser Ser Asn Thr Ala
      65             70             75             80

gcc acc aac gcc gac tgc gcc tgg ctg gag gcc cag gag gac gag gag      289
Ala Thr Asn Ala Asp Cys Ala Trp Leu Glu Ala Gln Glu Asp Glu Glu
      85             90             95

gtg ggc ttc ccc gtg agg ccc cag gtg ccc ctg agg ccc atg acc tac      337
Val Gly Phe Pro Val Arg Pro Gln Val Pro Leu Arg Pro Met Thr Tyr
      100            105            110

aag ggc gcc gtg gac ctg tcc cac ttc ctg aag gag aag ggc ggc ctg      385
Lys Gly Ala Val Asp Leu Ser His Phe Leu Lys Glu Lys Gly Gly Leu
      115            120            125

gag ggc ctg atc cac tcc cag aag agg cag gac atc ctg gac ctg tgg      433
Glu Gly Leu Ile His Ser Gln Lys Arg Gln Asp Ile Leu Asp Leu Trp
      130            135            140

gtg tac cac acc cag ggc tac ttc ccc gac tgg cag aac tac acc ccc      481
Val Tyr His Thr Gln Gly Tyr Phe Pro Asp Trp Gln Asn Tyr Thr Pro
      145            150            155            160

ggc ccc ggc atc agg ttc ccc ctg acc ttc ggc tgg tgc ttc aag ctg      529
Gly Pro Gly Ile Arg Phe Pro Leu Thr Phe Gly Trp Cys Phe Lys Leu
      165            170            175

gtg ccc gtg gag ccc gag aag gtg gag gag gcc aac gag ggc gag aac      577
Val Pro Val Glu Pro Glu Lys Val Glu Glu Ala Asn Glu Gly Glu Asn
      180            185            190

aac tgc ctg ctg cac ccc atg tcc cag cac ggc atc gag gac ccc gag      625
Asn Cys Leu Leu His Pro Met Ser Gln His Gly Ile Glu Asp Pro Glu
      195            200            205

aag gag gtg ctg gag tgg agg ttc gac tcc aag ctg gcc ttc cac cac      673

```

Lys Glu Val Leu Glu Trp Arg Phe Asp Ser Lys Leu Ala Phe His His
 210 215 220

gtg gcc agg gag ctg cac ccc gag tac tac aag gac tgc taa 715
 Val Ala Arg Glu Leu His Pro Glu Tyr Tyr Lys Asp Cys *
 225 230 235

agcc 719

<210> SEQ ID NO:4

<211> LENGTH: 237

<212> TYPE: PRT

<213> ORGANISM:Human Immunodeficiency Virus - 1

<400> SEQ ID NO:4

Met Asp Ala Met Lys Arg Gly Leu Cys Cys Val Leu Leu Leu Cys Gly
 1 5 10 15
 Ala Val Phe Val Ser Pro Ser Glu Ile Ser Ser Lys Arg Ser Val Pro
 20 25 30
 Gly Trp Ser Thr Val Arg Glu Arg Met Arg Arg Ala Glu Pro Ala Ala
 35 40 45
 Asp Arg Val Arg Arg Thr Glu Pro Ala Ala Val Gly Val Gly Ala Val
 50 55 60
 Ser Arg Asp Leu Glu Lys His Gly Ala Ile Thr Ser Ser Asn Thr Ala
 65 70 75 80
 Ala Thr Asn Ala Asp Cys Ala Trp Leu Glu Ala Gln Glu Asp Glu Glu
 85 90 95
 Val Gly Phe Pro Val Arg Pro Gln Val Pro Leu Arg Pro Met Thr Tyr
 100 105 110
 Lys Gly Ala Val Asp Leu Ser His Phe Leu Lys Glu Lys Gly Gly Leu
 115 120 125
 Glu Gly Leu Ile His Ser Gln Lys Arg Gln Asp Ile Leu Asp Leu Trp
 130 135 140
 Val Tyr His Thr Gln Gly Tyr Phe Pro Asp Trp Gln Asn Tyr Thr Pro
 145 150 155 160
 Gly Pro Gly Ile Arg Phe Pro Leu Thr Phe Gly Trp Cys Phe Lys Leu
 165 170 175
 Val Pro Val Glu Pro Glu Lys Val Glu Glu Ala Asn Glu Gly Glu Asn
 180 185 190
 Asn Cys Leu Leu His Pro Met Ser Gln His Gly Ile Glu Asp Pro Glu
 195 200 205
 Lys Glu Val Leu Glu Trp Arg Phe Asp Ser Lys Leu Ala Phe His His
 210 215 220
 Val Ala Arg Glu Leu His Pro Glu Tyr Tyr Lys Asp Cys
 225 230 235

<210> SEQ ID NO:5

<211> LENGTH: 671

<212> TYPE: DNA

<213> ORGANISM:Human Immunodeficiency Virus - 1

<220> FEATURE:

<221> NAME/KEY: CDS

<222> LOCATION: (12)...(662)

<400> SEQ ID NO:5

gatctgccac c atg gcc ggc aag tgg tcc aag agg tcc gtg ccc ggc tgg 50
 Met Ala Gly Lys Trp Ser Lys Arg Ser Val Pro Gly Trp
 1 5 10

tcc acc gtg agg gag agg atg agg agg gcc gag ccc gcc gcc gac agg 98

```

Ser Thr Val Arg Glu Arg Met Arg Arg Ala Glu Pro Ala Ala Asp Arg
  15                20                25

gtg agg agg acc gag ccc gcc gcc gtg ggc gtg ggc gcc gtg tcc agg      146
Val Arg Arg Thr Glu Pro Ala Ala Val Gly Val Gly Ala Val Ser Arg
  30                35                40                45

gac ctg gag aag cac ggc gcc atc acc tcc tcc aac acc gcc gcc acc      194
Asp Leu Glu Lys His Gly Ala Ile Thr Ser Ser Asn Thr Ala Ala Thr
      50                55                60

aac gcc gac tgc gcc tgg ctg gag gcc cag gag gac gag gag gtg ggc      242
Asn Ala Asp Cys Ala Trp Leu Glu Ala Gln Glu Asp Glu Glu Val Gly
      65                70                75

ttc ccc gtg agg ccc cag gtg ccc ctg agg ccc atg acc tac aag ggc      290
Phe Pro Val Arg Pro Gln Val Pro Leu Arg Pro Met Thr Tyr Lys Gly
      80                85                90

gcc gtg gac ctg tcc cac ttc ctg aag gag aag ggc ggc ctg gag ggc      338
Ala Val Asp Leu Ser His Phe Leu Lys Glu Lys Gly Gly Leu Glu Gly
      95                100                105

ctg atc cac tcc cag aag agg cag gac atc ctg gac ctg tgg gtg tac      386
Leu Ile His Ser Gln Lys Arg Gln Asp Ile Leu Asp Leu Trp Val Tyr
     110                115                120                125

cac acc cag ggc tac ttc ccc gac tgg cag aac tac acc ccc ggc ccc      434
His Thr Gln Gly Tyr Phe Pro Asp Trp Gln Asn Tyr Thr Pro Gly Pro
      130                135                140

ggc atc agg ttc ccc ctg acc ttc ggc tgg tgc ttc aag ctg gtg ccc      482
Gly Ile Arg Phe Pro Leu Thr Phe Gly Trp Cys Phe Lys Leu Val Pro
      145                150                155

gtg gag ccc gag aag gtg gag gag gcc aac gag ggc gag aac aac tgc      530
Val Glu Pro Glu Lys Val Glu Glu Ala Asn Glu Gly Glu Asn Asn Cys
      160                165                170

gcc gcc cac ccc atg tcc cag cac ggc atc gag gac ccc gag aag gag      578
Ala Ala His Pro Met Ser Gln His Gly Ile Glu Asp Pro Glu Lys Glu
      175                180                185

gtg ctg gag tgg agg ttc gac tcc aag ctg gcc ttc cac cac gtg gcc      626
Val Leu Glu Trp Arg Phe Asp Ser Lys Leu Ala Phe His His Val Ala
     190                195                200                205

agg gag ctg cac ccc gag tac tac aag gac tgc taa agcccgggc      671
Arg Glu Leu His Pro Glu Tyr Tyr Lys Asp Cys *
      210                215

```

<210> SEQ ID NO:6

<211> LENGTH: 217

<212> TYPE: PRT

<213> ORGANISM: Human Immunodeficiency Virus - 1

<400> SEQ ID NO:6

Met Ala Gly Lys Trp Ser Lys Arg Ser Val Pro Gly Trp Ser Thr Val

```

1           5           10           15
Arg Glu Arg Met Arg Arg Ala Glu Pro Ala Ala Asp Arg Val Arg Arg
20           25           30
Thr Glu Pro Ala Ala Val Gly Val Gly Ala Val Ser Arg Asp Leu Glu
35           40           45
Lys His Gly Ala Ile Thr Ser Ser Asn Thr Ala Ala Thr Asn Ala Asp
50           55           60
Cys Ala Trp Leu Glu Ala Gln Glu Asp Glu Glu Val Gly Phe Pro Val
65           70           75           80
Arg Pro Gln Val Pro Leu Arg Pro Met Thr Tyr Lys Gly Ala Val Asp
85           90           95
Leu Ser His Phe Leu Lys Glu Lys Gly Gly Leu Glu Gly Leu Ile His
100          105          110
Ser Gln Lys Arg Gln Asp Ile Leu Asp Leu Trp Val Tyr His Thr Gln
115          120          125
Gly Tyr Phe Pro Asp Trp Gln Asn Tyr Thr Pro Gly Pro Gly Ile Arg
130          135          140
Phe Pro Leu Thr Phe Gly Trp Cys Phe Lys Leu Val Pro Val Glu Pro
145          150          155          160
Glu Lys Val Glu Glu Ala Asn Glu Gly Glu Asn Asn Cys Ala Ala His
165          170          175
Pro Met Ser Gln His Gly Ile Glu Asp Pro Glu Lys Glu Val Leu Glu
180          185          190
Trp Arg Phe Asp Ser Lys Leu Ala Phe His His Val Ala Arg Glu Leu
195          200          205
His Pro Glu Tyr Tyr Lys Asp Cys Ser
210          215

```

<210> SEQ ID NO:7

<211> LENGTH: 720

<212> TYPE: DNA

<213> ORGANISM: Human Immunodeficiency Virus - 1

<220> FEATURE:

<221> NAME/KEY: CDS

<222> LOCATION: (2)...(715)

<400> SEQ ID NO:7

```

c atg gat gca atg aag aga ggg ctc tgc tgt gtg ctg ctg ctg tgt gga      49
  Met Asp Ala Met Lys Arg Gly Leu Cys Cys Val Leu Leu Leu Cys Gly
    1           5           10           15

gca gtc ttc gtt tcg ccc agc gag atc tcc tcc aag agg tcc gtg ccc      97
Ala Val Phe Val Ser Pro Ser Glu Ile Ser Ser Lys Arg Ser Val Pro
    20           25           30

ggc tgg tcc acc gtg agg gag agg atg agg agg gcc gag ccc gcc gcc     145
Gly Trp Ser Thr Val Arg Glu Arg Met Arg Arg Ala Glu Pro Ala Ala
    35           40           45

gac agg gtg agg agg acc gag ccc gcc gcc gtg ggc gtg ggc gcc gtg     193
Asp Arg Val Arg Arg Thr Glu Pro Ala Ala Val Gly Val Gly Ala Val
    50           55           60

tcc agg gac ctg gag aag cac ggc gcc atc acc tcc tcc aac acc gcc     241
Ser Arg Asp Leu Glu Lys His Gly Ala Ile Thr Ser Ser Asn Thr Ala
    65           70           75           80

gcc acc aac gcc gac tgc gcc tgg ctg gag gcc cag gag gac gag gag     289
Ala Thr Asn Ala Asp Cys Ala Trp Leu Glu Ala Gln Glu Asp Glu Glu

```


	85	90	95	
gtg ggc ttc ccc gtg agg ccc cag gtg ccc ctg agg ccc atg acc tac				337
Val Gly Phe Pro Val Arg Pro Gln Val Pro Leu Arg Pro Met Thr Tyr				
	100	105	110	
aag ggc gcc gtg gac ctg tcc cac ttc ctg aag gag aag ggc ggc ctg				385
Lys Gly Ala Val Asp Leu Ser His Phe Leu Lys Glu Lys Gly Gly Leu				
	115	120	125	
gag ggc ctg atc cac tcc cag aag agg cag gac atc ctg gac ctg tgg				433
Glu Gly Leu Ile His Ser Gln Lys Arg Gln Asp Ile Leu Asp Leu Trp				
	130	135	140	
gtg tac cac acc cag ggc tac ttc ccc gac tgg cag aac tac acc ccc				481
Val Tyr His Thr Gln Gly Tyr Phe Pro Asp Trp Gln Asn Tyr Thr Pro				
	145	150	155	160
ggc ccc ggc atc agg ttc ccc ctg acc ttc ggc tgg tgc ttc aag ctg				529
Gly Pro Gly Ile Arg Phe Pro Leu Thr Phe Gly Trp Cys Phe Lys Leu				
	165	170	175	
gtg ccc gtg gag ccc gag aag gtg gag gag gcc aac gag ggc gag aac				577
Val Pro Val Glu Pro Glu Lys Val Glu Glu Ala Asn Glu Gly Glu Asn				
	180	185	190	
aac tgc gcc gcc cac ccc atg tcc cag cac ggc atc gag gac ccc gag				625
Asn Cys Ala Ala His Pro Met Ser Gln His Gly Ile Glu Asp Pro Glu				
	195	200	205	
aag gag gtg ctg gag tgg agg ttc gac tcc aag ctg gcc ttc cac cac				673
Lys Glu Val Leu Glu Trp Arg Phe Asp Ser Lys Leu Ala Phe His His				
	210	215	220	
gtg gcc agg gag ctg cac ccc gag tac tac aag gac tgc taa				715
Val Ala Arg Glu Leu His Pro Glu Tyr Tyr Lys Asp Cys *				
	225	230	235	
agccc				720

<210> SEQ ID NO:8

<211> LENGTH: 237

<212> TYPE: PRT

<213> ORGANISM:Human Immunodeficiency Virus - 1

<400> SEQ ID NO:8

Met Asp Ala Met Lys Arg Gly Leu Cys Cys Val Leu Leu Leu Cys Gly				
1	5	10	15	
Ala Val Phe Val Ser Pro Ser Glu Ile Ser Ser Lys Arg Ser Val Pro				
	20	25	30	
Gly Trp Ser Thr Val Arg Glu Arg Met Arg Arg Ala Glu Pro Ala Ala				
	35	40	45	
Asp Arg Val Arg Arg Thr Glu Pro Ala Ala Val Gly Val Gly Ala Val				
	50	55	60	
Ser Arg Asp Leu Glu Lys His Gly Ala Ile Thr Ser Ser Asn Thr Ala				
	65	70	75	80
Ala Thr Asn Ala Asp Cys Ala Trp Leu Glu Ala Gln Glu Asp Glu Glu				
	85	90	95	
Val Gly Phe Pro Val Arg Pro Gln Val Pro Leu Arg Pro Met Thr Tyr				
	100	105	110	

Lys Gly Ala Val Asp Leu Ser His Phe Leu Lys Glu Lys Gly Gly Leu
 115 120 125
 Glu Gly Leu Ile His Ser Gln Lys Arg Gln Asp Ile Leu Asp Leu Trp
 130 135 140
 Val Tyr His Thr Gln Gly Tyr Phe Pro Asp Trp Gln Asn Tyr Thr Pro
 145 150 155 160
 Gly Pro Gly Ile Arg Phe Pro Leu Thr Phe Gly Trp Cys Phe Lys Leu
 165 170 175
 Val Pro Val Glu Pro Glu Lys Val Glu Glu Ala Asn Glu Gly Glu Asn
 180 185 190
 Asn Cys Ala Ala His Pro Met Ser Gln His Gly Ile Glu Asp Pro Glu
 195 200 205
 Lys Glu Val Leu Glu Trp Arg Phe Asp Ser Lys Leu Ala Phe His His
 210 215 220
 Val Ala Arg Glu Leu His Pro Glu Tyr Tyr Lys Asp Cys
 225 230 235

<210> SEQ ID NO:9

<211> LENGTH: 4945

<212> TYPE: DNA

<213> ORGANISM: E. coli

<400> SEQ ID NO:9

tcgcgcgttt	cggatgatgac	ggtgaaaacc	tctgacacat	gcagctcccg	gagacggtca	60
cagcttgtct	gtaagcggat	gccgggagca	gacaagcccg	tcagggcgcg	tcagcgggtg	120
ttggcggtg	tcggggctgg	cttaactatg	cggcatcaga	gcagattgta	ctgagagtgc	180
accatatgcy	gtgtgaaata	cgcacacagat	gcgtaaggag	aaaataccgc	atcagattgg	240
ctattggcca	ttgcatacgt	tgtatccata	tcataatatg	tacatttata	ttggctcatg	300
tccaacatta	ccgcatgtt	gacattgatt	attgactagt	tattaatagt	aatcaattac	360
ggggtcatta	gttcatagcc	catatatgga	gttcgcggtt	acataactta	cggtaaatgg	420
ccgcctggc	tgaccgcca	acgaccccg	ccattgacg	tcaataatga	cgtatgttcc	480
catagtaacg	ccaataggga	ctttccattg	acgtcaatgg	gtggagtatt	tacggtaaac	540
tgcccacttg	gcagtacatc	aagtgtatca	tatgccaatg	acgcccccta	ttgacgtcaa	600
tgacggtaaa	tggcccgcct	ggcattatgc	ccagtacatg	accttatggg	actttcctac	660
ttggcagtac	atctacgtat	tagtcacgcg	tattaccatg	gtgatgcggg	tttggcagta	720
catcaatggg	cgtggatagc	ggtttgactc	acggggattt	ccaagtctcc	accccatgta	780
cgtcaatggg	agtttgtttt	ggcaccaaaa	tcaacgggac	tttccaaaat	gtcgtaacaa	840
ctccgcccc	ttgacgcaaa	tgggcggtag	gcgtgtacgg	tgggagggtc	atataagcag	900
agtcgttta	gtgaaccgtc	agatcgctg	gagacgccat	ccacgctgtt	ttgacctcca	960
tagaagacac	cgggaccgat	ccagcctccg	cggccgggaa	cgggtgcattg	gaacgcggat	1020
tccccgtgcc	aagagtgcag	taagtaccgc	ctatagactc	tataggcaca	cccctttggc	1080
tcttatgcat	gctatactgt	ttttggcttg	gggcctatac	acccccgctt	ccttatgcta	1140
taggtgatgg	tataacttag	cctatagggtg	tgggttattg	accattattg	accactcccc	1200
tattgggtgac	gatactttcc	attactaatc	cataacatgg	ctctttgcca	caactatctc	1260
tattggctat	atgccaatac	tctgtccttc	agagactgac	acggactctg	tattttttaca	1320
ggatgggggtc	ccatttatta	tttaccaaat	cacatatata	acaacgcctg	cccccggtgc	1380
cgcagttttt	attaacata	gcgtgggatc	tccacgcgaa	tctcgggtac	gtgttccgga	1440
catgggctct	tctccggtag	cggcgaggct	tccacatccg	agccctgggtc	ccatgcctcc	1500
agcggctcat	ggtcgctcgg	cagtccttg	ctcctaacag	tggaggccag	acttaggcac	1560
agcacaatgc	ccaccaccac	cagtgtgccg	cacaaggccg	tggcggtagg	gtatgtgtct	1620
gaaaatgagc	gtggagattg	ggctcgcacg	gctgacgcag	atggaagact	taaggcagcg	1680
gcagaagaag	atgcaggcag	ctgagttggt	gtattctgat	aagagtcaga	ggtaactccc	1740
gttgcggtgc	tgtaaacggt	ggagggcagt	gtagtctgag	cagtactcgt	tgctgcccg	1800
cgcgccacca	gacataatag	ctgacagact	aacagactgt	tcctttccat	gggtcttttc	1860
tgagtcacc	gtccttagat	caccatggat	gcaatgaaga	gagggctctg	ctgtgtgctg	1920
ctgctgtgtg	gagcagctct	cgtttcgccc	agcgaatctc	gctgtgcctt	ctagttgcc	1980
gccatctgtt	gtttgcccc	cccccggtgc	ttccttgacc	ctggaagggtg	ccactcccac	2040
tgctctttcc	taataaaatg	aggaaattgc	atcgcatgtg	ctgagtaggt	gtcattctat	2100
tctggggggg	ggggtggggc	aggacagcaa	gggggaggat	tgggaagaca	atagcaggca	2160
tgctggggat	gcggtgggct	ctatggcccg	tgcggccagg	tgctgaagaa	ttgacccggt	2220
tcctcctggg	ccagaaagaa	gcaggcacat	ccccttctct	gtgacacacc	ctgtccacgc	2280

```

ccctgggttct tagttccagc cccactcata ggacactcat agctcaggag ggctccgcct 2340
tcaatccac ccgctaaagt acttggagcg gtctctccct ccctcatcag cccaccaaac 2400
caaacctagc ctccaagagt gggaagaaat taaagcaaga taggctatta agtgcagagg 2460
gagagaaaat gcctccaaca tgtgaggaag taatgagaga aatcatagaa tttcttccgc 2520
ttcctcgctc actgactcgc tgcgctcggg cgttcggctg cggcgagcgg tatcagctca 2580
ctcaaaggcg gtaatacggg tatccacaga atcaggggat aacgcaggaa agaacatgtg 2640
agcaaaaggc cagcaaaagg ccaggaaccg taaaaaggcc gcgttgctgg cgtttttcca 2700
taggtccgc cccctgacg agcatcaca aaatcgacgc tcaagtcaga ggtggcgaaa 2760
cccgacagga ctataaagat accaggcgtt tccccctgga agctccctcg tgcgctctcc 2820
tgttccgacc ctgccgtta ccggatacct gtccgccttt ctcccttcgg gaagcgtggc 2880
gctttctcat agctcacgt gtaggtatct cagttcggtg taggtcgttc gctccaagct 2940
gggctgtgtg cacgaacccc ccgttcagcc cgaccgctgc gccttatccg gtaactatcg 3000
tcttgagtc aacccggtaa gacacgactt atcgccactg gcagcagcca ctggtaacag 3060
gattagcaga gcgaggtatg taggcggtgc tacagatttc ttgaagtggg ggcctaacta 3120
cggctacact agaagaacag tatttggat ctgcgctctg ctgaagccag ttaccttcgg 3180
aaaaagagtt ggtagctctt gatccggcaa acaaaccacc gctggtagcg gtgggttttt 3240
tgtttgcaag cagcagatta cgcgcagaaa aaaaggatct caagaagatc ctttgatctt 3300
ttctacgggg tctgacgctc agtggaaacga aaactcacgt taagggattt tggatcatgag 3360
attatcaaaa aggatcttca cctagatcct tttaaattaa aaatgaagt tttaaataat 3420
ctaaagtata tatgagtaaa cttggtctga cagttacca tgcttaata gtgaggcacc 3480
tatctcagcg atctgtctat ttcgttcac tcagattgcc tgactcgggg gggggggggc 3540
ctgaggtctg cctcgtgaag aaggtgttgc tgactcatac caggcctgaa tcgccccatc 3600
atccagccag aaagtgaggg agccacggtt gatgagagct ttgtttagg tggaccagtt 3660
ggtgattttg aacttttgc ttgccacgga acggtctgcg ttgtcgggaa gatgcgtgat 3720
ctgatccctc aactcagcaa aagttcgatt tattcaacaa agccgcctgc cgtcaagtc 3780
agcgtaatgc tctgccagt ttacaaccaa ttaaccaatt ctgattagaa aaactcatcg 3840
agcataaat gaaactgcaa tttattcata tcaggattat caataccata tttttgaaaa 3900
agccgtttct gtaatgaagg agaaaactca ccgaggcagt tccataggat ggcaagatcc 3960
tggtatcggg ctgcgattcc gactcgtcca acatcaatac aacctattaa tttccctcg 4020
tcaaaaataa ggttatcaag tgagaaatca ccatgagtga cgactgaatc cggtgagaat 4080
ggcaaaagct tatgcatctt tttccagact tgttcaacag gccagccatt acgctcgtca 4140
tcaaaatcac tcgcatcaac caaacggtta ttcattcgtg attgcgcctg agcgagacga 4200
aatacgcgat cgctgttaaa aggacaatta caaacaggaa tcgaatgcaa cgggcgcagg 4260
aacactgcca gcgcataac aatattttca cctaatacag gatattcttc taatacctgg 4320
aatgctgttt tcccggggat cgcagtggg agtaaccatg catcatcagg agtacggata 4380
aaatgcttga tggcgggaag aggcataaat tccgtcagcc agtttagtct gaccatctca 4440
tctgtaacat cattggcaac gctacctttg ccatgtttca gaaacaactc tggcgcacgc 4500
ggcttcccat acaatcgata gattgtcgca cctgattgcc cgacattatc gcgagcccat 4560
ttatacccat ataaatcagc atccatgttg gaatttaatc gcggcctcga gcaagacgtt 4620
tcccgttgaa tatggctcat aacaccctt gtattactgt ttatgtaagc agacagtttt 4680
attgttcatt atgatatatt tttatcttgt gcaatgtaac atcagagatt ttgagacaca 4740
acgtgggttt ccccccccc ccattattga agcatttatc agggttattg tctcatgagc 4800
ggatacatat ttgaatgtat ttagaaaaat aaacaaatag gggttccgcg cacatttccc 4860
cgaaaagtgc cacctgacgt ctaagaaacc attattatca tgacattaac ctataaaaa 4920
aggcgtatca cgaggccctt tcgtc 4945

```

<210> SEQ ID NO:10

<211> LENGTH: 23

<212> TYPE: DNA

<213> ORGANISM:Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: oligonucleotide

<400> SEQ ID NO:10

ctatataagc agagctcgtt tag

23

<210> SEQ ID NO:11

<211> LENGTH: 30

<212> TYPE: DNA
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: oligonucleotide

<400> SEQ ID NO:11
 gtagcaaaga tctaaggacg gtgactgcag

30

<210> SEQ ID NO:12
 <211> LENGTH: 39
 <212> TYPE: DNA
 <213> ORGANISM: Artificial Sequence

<220> FEATURE:
 <223> OTHER INFORMATION: oligonucleotide

<400> SEQ ID NO:12
 gtatgtgtct gaaaatgagc gtggagattg ggctcgcac

39

<210> SEQ ID NO:13
 <211> LENGTH: 39
 <212> TYPE: DNA
 <213> ORGANISM: Artificial Sequence

<220> FEATURE:
 <223> OTHER INFORMATION: oligonucleotide

<400> SEQ ID NO:13
 gtgcgagccc aatctccacg ctcattttca gacacatac

39

<210> SEQ ID NO:14
 <211> LENGTH: 4432
 <212> TYPE: DNA
 <213> ORGANISM: E. coli

<400> SEQ ID NO:14
 tcgcgcgttt cgggtgatgac ggtgaaaacc tctgacacat gcagctcccg gagacgggtca 60
 cagcttgtct gtaagcggat gccgggagca gacaagcccg tcagggcgcg tcagcgggtg 120
 ttggcggtg tcggggctgg cttaactatg cggcatcaga gcagattgta ctgagagtgc 180
 accatatgcg gtgtgaaata ccgcacagat gcgtaaggag aaaataccgc atcagattgg 240
 ctattggcca ttgcatacgt tgtatccata tcataatatg tacatttata ttggctcatg 300
 tccaacatta ccgccatgtt gacattgatt attgactagt tattaatagt aatcaattac 360
 ggggtcatta gttcatagcc catatatgga gttccgcgtt acataactta cggtaaatgg 420
 ccgcctggc tgaccgcca acgacccccg ccattgacg tcaataatga cgtatgttcc 480
 catagtaacg ccaataggga ctttccattg acgtcaatgg gtggagtatt tacggtaaac 540
 tgcccacttg gcagtacatc aagtgtatca tatgccagt acgcccccta ttgacgtcaa 600
 tgacggtaaa tggcccgcct ggcattatgc ccagtacatg accttatggg actttcctac 660
 ttggcagtac atctacgtat tagtcacgc tattaccatg gtgatgcgtt tttggcagta 720
 catcaatggg cgtggatagc ggtttgactc acggggattt ccaagtctcc accccattga 780
 cgtcaatggg agtttgtttt ggcacaaaaa tcaacgggac tttccaaaat gtcgtaacaa 840
 ctccgcccc a ttgacgcaa tgggcggtag gcgtgtacgg tgggaggtct atataagcag 900
 agctcgttta gtgaaccgtc agatcgctg gagacgccat ccacgtgtt ttgacctcca 960
 tagaagacac cgggaccgat ccagcctccg cggccgggaa cgggtgcattg gaacgcggat 1020
 tccccgtgcc aagagtgcg taagtaccgc ctatagagtc tataggccca cccccttggc 1080
 ttttatgca tgctatactg tttttggctt ggggtctata ccccccgct tcctcatgtt 1140
 atagggtatg gtatagctta gcctatagg gtgggttatt gaccattatt gaccactccc 1200
 ctattgggtga cgatactttc cattaactaa ccataacatg gctctttgcc acaactctct 1260
 ttattggcta tatgccaaata cactgtcctt cagagactga cacggactct gtatttttac 1320
 aggatggggc ctcattttatt atttacaaat tcacatatac aacaccaccg tccccagtcg 1380
 ccgcagtttt tattaacat aacgtgggat ctccacgcga atctcgggta cgtgttccgg 1440

acatgggctc	ttctccggtg	gcgggcgagc	ttctacatcc	gagccctgct	cccatgcctc	1500
cagcgactca	tggtcgctcg	gcagctcctt	gctcctaaca	gtggaggcca	gacttaggca	1560
cagcacgatg	cccaccacca	ccagtggtcc	gcacaaggcc	gtggcggtag	ggtatgtgtc	1620
tgaaaatgag	ctcgggggagc	gggcttgac	cgtgacgca	tttggagac	ttaaggcagc	1680
ggcagaagaa	gatgcaggca	gctgagttgt	tgtgttctga	taagagtcag	aggtaactcc	1740
cggtgctgtg	ctgttaacgg	tggaggggcag	tgtagtctga	gcagtactcg	ttgctgccgc	1800
gcgcgccacc	agacataata	gctgacagac	taacagactg	ttcctttcca	tgggtctttt	1860
ctgcagtcac	cgctccttaga	tctgctgtgc	cttctagttg	ccagccatct	gttgtttgcc	1920
cctccccgt	gccttccttg	accctggaag	gtgccactcc	cactgtcctt	tcctaataaaa	1980
atgaggaat	tgcacgcac	tgtctgagta	ggtgtcattc	tattctgggg	ggtgggggtg	2040
ggcagcacag	caagggggag	gattgggaag	acaatagcag	gcatgctggg	gatgcgggtg	2100
gctctatggg	taccagggtg	ctgaagaatt	gacccggttc	ctcctggggc	agaaagaagc	2160
aggcacatcc	ccttctctgt	gacacaccct	gtccacgccc	ctgggtctta	gttccagccc	2220
cactcatagg	acactcatag	ctcaggaggg	ctccgccttc	aatcccaccc	gctaaagtac	2280
ttggagcggt	ctctccctcc	ctcatcagcc	caccaaacca	aacctagcct	ccaagagtgg	2340
gaagaaatta	aagcaagata	ggctattaag	tgcagaggga	gagaaaatgc	ctccaacatg	2400
tgaggaaagta	atgagagaaa	tcatagaatt	tcttcgcgtt	cctcgctcac	tgactcgctg	2460
cgctcggtcg	ttcggtcgcg	gcgagcggtg	tcagctcact	caaaggcggt	aatacggtta	2520
tccacagaat	caggggataa	cgcaggaaa	aacatgtgag	caaaaggcca	gcaaaaggcc	2580
aggaaaccgtg	aaaaggccgc	gttgcgtggc	tttttccata	ggctccgccc	ccctgacgag	2640
catcacaaaa	atcgacgctc	aagtcagagg	tggcgaaacc	cgacaggact	ataaagatac	2700
caggcgtttc	cccctggaag	ctccctcgtg	cgtctcctcg	ttccgaccct	gccgcttacc	2760
ggatacctgt	cgcctttct	cccttcggga	agcgtggcgc	tttctcaatg	ctcacgctgt	2820
aggtatctca	gttcggtgta	ggtcgttcgc	tccaagctgg	gctgtgtgca	cgaaccccc	2880
gttcagcccg	accgctgctc	cttatccggt	aactatcgtc	ttgagtccaa	cccggtaaga	2940
cacgacttat	cgccactggc	agcagccact	ggtaacagga	ttagcagagc	gaggtatgta	3000
ggcggtgtcta	cagagttctt	gaagtgggtg	cctaactacg	gctacactag	aaggacagta	3060
tttggtatct	gcgctctgct	gaagccagtt	accttcggaa	aaagagttgg	tagctcttga	3120
tccggcaaac	aaaccaccgc	tggtagcggt	ggtttttttg	tttgcaagca	gcagattacg	3180
cgcagaaaaa	aaggatctca	agaagatcct	ttgatctttt	ctacggggtc	tgacgctcag	3240
tggaaacgaaa	actcacgtta	agggattttg	gtcatgagat	tatcaaaaag	gatcttcacc	3300
tagatccttt	taaattaaaa	atgaagtttt	aatcaatct	aaagtatata	tgagtaaact	3360
tggtctgaca	gttaccaatg	cttaatcagt	gaggcaccta	tctcagcgat	ctgtctattt	3420
cgttcatcca	tagttgcctg	actccccgtc	gtgtagataa	ctacgatacg	ggagggctta	3480
ccatctggcc	ccagtgtctg	aatgataccg	cgagaccac	gctcaccggc	tccagattta	3540
tcagcaataa	accagccagc	cgggaaggcc	gagcgcagaa	gtggtcctgc	aactttatcc	3600
gcctccatcc	agtctattaa	ttgttgccgg	gaagctagag	taagtagttc	gccagttaat	3660
agtttgcgca	acgttggtgc	cattgctaca	ggcatcgtgg	tgacagctc	gtcgtttggt	3720
atggcttcat	tcagctccgg	ttcccaacga	tcaaggcgag	ttacatgatc	ccccatgttg	3780
tgcaaaaaag	cgtttagctc	cttcggctct	ccgatcgttg	tcagaagtaa	gttggccgca	3840
gtgttatcac	ctatggttat	ggcagcactg	cataattctc	ttactgtcat	gccatccgta	3900
agatgctttt	ctgtgactgg	tgagtactca	accaagtcac	tctgagaata	gtgtatgcgg	3960
cgaccgagtt	gctcttgccc	ggcgtcaata	cgggataata	ccgcgccaca	tagcagaact	4020
ttaaaagtgc	tcatcattgg	aaaacgttct	tggggcgcaa	aactctcaag	gatcttaccg	4080
ctgttgagat	ccagttcgat	gtaaccact	cgtgcaccca	actgatcttc	agcatctttt	4140
actttcacca	gcgtttctgg	gtgagcaaaa	acaggaaggc	aaaatgccgc	aaaaaaggga	4200
ataaggcgca	cacggaaatg	ttgaatactc	atactcttcc	tttttcaata	ttattgaagc	4260
atttatcagg	gttattgtct	catgagcgga	tacatatttg	aatgtattta	gaaaaataaa	4320
caaatagggg	ttccgcgcac	atttccccga	aaagtgccac	ctgacgtcta	agaaaccatt	4380
attatcatga	cattaaccta	taaaaatagg	cgtatcacga	ggccctttcg	tc	4432

<210> SEQ ID NO:15

<211> LENGTH: 4864

<212> TYPE: DNA

<213> ORGANISM: E. coli

<400> SEQ ID NO:15

tcgcgcgttt	cggtgatgac	ggtgaaaacc	tctgacacat	gcagctcccc	gagacgggtca	60
cagcttgtct	gtaagcggat	gccggggagca	gacaagcccc	tcagggcgcg	tcagcgggtg	120
ttggcggttg	tcggggctgg	cttaactatg	cgcatcaga	gcagattgta	ctgagagtgc	180
accatattgcg	gtgtgaaata	ccgcacagat	gcgtaaggag	aaaaatccgc	atcagattgg	240

ctattggcca	ttgcatacgt	tgtatccata	tcataatatg	tacatttata	ttggctcatg	300
tccaacatta	ccgccatggt	gacattgatt	attgactagt	tattaatagt	aatcaattac	360
ggggtcatta	gttcatagcc	catatatgga	gttccgcgtt	acataactta	cggtaaatgg	420
cccgctggc	tgaccgcca	acgacccccg	cccatgacg	tcaataatga	cgtatgttcc	480
catagtaacg	ccaataggga	ctttccattg	acgtcaatgg	gtggagtatt	tacggtaaac	540
tgcccacttg	gcagtacatc	aagtgtatca	tatgccaatg	acgcccccta	ttgacgtcaa	600
tgacggtaaa	tggcccgccct	ggcattatgc	ccagtacatg	accttatggg	actttcctac	660
ttggcagtac	atctacgtat	tagtcatcgc	tattaccatg	gtgatgcggg	tttggcagta	720
catcaatggg	cgtggatagc	ggtttgactc	acggggattt	ccaagtctcc	accccatgta	780
cgtcaatggg	agtttgtttt	ggcaccaaaa	tcaacgggac	tttccaaaat	gtcgtaacaa	840
ctccgcccc	ttgacgcaaa	tgggcggtag	gcgtgtacgg	tgggagggtc	atataagcag	900
agctcgttta	gtgaaccgtc	agatcgccctg	gagacgccat	ccacgctgtt	ttgacctcca	960
tagaagacac	cgggaccgat	ccagcctccg	cggccgggaa	cgggtgcattg	gaacgcggat	1020
tccccgtgcc	aagagtgcag	taagtaccgc	ctatagagtc	tataggccca	cccccttggc	1080
ttcttatgca	tgtatactg	tttttggctt	ggggtctata	cacccccgct	tcctcatgtt	1140
ataggtagtg	gtatagctta	gcctataggt	gtgggttatt	gaccattatt	gaccactccc	1200
ctatttggtga	cgatactttc	cattactaat	gcataacatg	gctctttgcc	acaactctct	1260
ttattggcta	tatgccaaata	cactgtcctt	cagagactga	cacggactct	gtattttttac	1320
aggatgggggt	ctcatatttatt	atttacaat	tcacatatac	aacaccaccg	tccccagtgc	1380
ccgcagtttt	tattaaacat	aacgtgggat	ctccacgcga	atctcgggta	cgtgttccgg	1440
acatgggctc	ttctccggta	gcggcgggagc	ttctacatcc	gagccctgct	cccatgcctc	1500
cagcgactac	tggctcgctcg	gcagctcctt	gtccctaaca	gtggaggcca	gacttaggca	1560
cagcacgatg	cccaccacca	ccagtgtgcc	gcacaaggcc	gtggcggtag	ggtagtgttc	1620
tgaatagag	ctcggggagc	gggcttgac	gcgtgacgca	tttggaagac	ttaaggcagc	1680
ggcagaagaa	gatgcaggca	gctgagttgt	tgtgttctga	taagagtcag	aggtaactcc	1740
cgttgccgtg	ctgttaacgg	tggagggcag	tgtagtctga	gcagtactcg	ttgctgccgc	1800
gcgcgccacc	agacataata	gctgacagac	taacagactg	ttcctttcca	tgggtctttt	1860
ctgcagtcac	cgctcttaga	tctgctgtgc	cttctagtgt	ccagccatct	gttgttttgc	1920
cctccccgtg	gccttccctg	accctggaag	gtgccactcc	cactgtcctt	tcctaataaa	1980
atgaggaaat	tgcctcgcat	tgtctagta	gggtctatct	tattctgggg	gggtgggtgg	2040
ggcagcacag	caagggggag	gattgggaag	acaatagcag	gcattgctggg	gatgcgggtg	2100
gctctatggg	taccaggtg	ctgaagaatt	gacccgggtc	ctcctggggc	agaaagaagc	2160
aggcacatcc	ccttctctgt	gacacaccct	gtccacgccc	ctggttctta	gttccagccc	2220
cactcatagg	acactcatag	ctcaggaggg	ctccgccttc	aatcccaccc	gctaaagtac	2280
ttggagcggg	ctctccctcc	ctcatcagcc	caccaaacca	aacctagcct	ccaagagtgg	2340
gaagaaatta	aagcaagata	ggctattaag	tgcagaggga	gagaaaaatgc	ctccaacatg	2400
tgaggaaagta	atgagagaaa	tcatagaatt	tcttccgctt	cctcgctcac	tgactcgctg	2460
cgctcggtcg	ttcggtcg	gcgagcggta	tcagctcaact	caaaggcggg	aatacggtta	2520
tccacagaat	caggggataa	cgcaggaaag	aacatgtgag	caaaaaggcca	gcaaaaggcc	2580
aggaaccgta	aaaaggccgc	gttgctggcg	ttttccata	ggctccgccc	ccctgacgag	2640
catcacaaaa	atcgacgctc	aagtcagagg	tggcgaaacc	cgacaggact	ataaagatac	2700
caggcggttc	ccccggaag	ctccctcggt	cgctctcctg	ttccgaccct	gccgcttacc	2760
ggatacctgt	ccgcctttct	cccttcggga	agcgtggcgc	tttctcaatg	ctcacgctgt	2820
aggatctca	gttcgggtga	ggctgttcgc	tccaagctgg	gctgtgtgca	cgaaccccc	2880
gttcagccc	accgctgcgc	cttatccggt	aactatcgtc	ttgagtccaa	cccggtaaga	2940
cacgacttat	cgccactggc	agcagccact	ggtaacagga	ttagcagagc	gaggtatgta	3000
ggcgggtgcta	cagagttctt	gaagtgggtg	cctaactacg	gctacactag	aaggacagta	3060
tttggtatct	gcgctctgct	gaagccagtt	accttcggaa	aaagagtgg	tagctcttga	3120
tccggcaaac	aaaccaccgc	tggtagcggg	ggtttttttg	tttgcaagca	gcagattacg	3180
cgcagaaaaa	aaggatctca	agaagatcct	ttgatctttt	ctacgggggc	tgacgctcag	3240
tggaaacgaaa	actcacgtta	agggattttg	gtcatgagat	tatcaaaaag	gatcttcacc	3300
tagatccttt	taaataaaaa	atgaagtttt	aaatcaatct	aaagtatata	tgagtaaact	3360
tggcttgaca	gttaccaatg	cttaatcagt	gaggcaccta	tctcagcgat	ctgtctattt	3420
cgttcatcca	tagttgectg	actccggggg	ggggggggcg	tgaggtctgc	ctcgtgaaga	3480
agggtgtgct	gactcatacc	aggcctgaat	cgccccatca	tccagccaga	aagtgagggg	3540
gccacgggtg	atgagagctt	tgttgtaggt	ggaccagttg	gtgattttga	acttttgctt	3600
tgccacggaa	cggctctcgt	tgtcgggaag	atgcgtgatc	tgatccttca	actcagcaaa	3660
agttcgattt	attcaacaaa	gccgcggtcc	cgtaagtcga	gcgtaatgct	ctgccagtg	3720
tacaaccaat	taaccaattc	tgattagaaa	aactcatcga	gcatcaaatg	aaactgcaat	3780
ttattcatat	caggattatc	aataccatat	ttttgaaaaa	gccgtttctg	taatgaagga	3840
gaaaactcac	cgaggcagtt	ccataggatg	gcaagatcct	ggtatcggtc	tgcgattccc	3900

actcgtccaa	catcaataca	acctattaat	ttccccctcgt	caaaaataag	gttatcaagt	3960
gagaaatcac	catgagtac	gactgaatcc	ggtgagaatg	gcaaaagcct	atgcatttct	4020
ttccagactt	gttcaacagg	ccagccatta	cgctcgtcat	caaaatcact	cgcatcaacc	4080
aaaccggtat	tcattcgtga	ttgcgcctga	gcgagacgaa	atacgcgatc	gctgttaaaa	4140
ggacaattac	aaacaggaat	cgaatgcaac	cggcgacgga	acactgccag	cgcatcaaca	4200
atattttcac	ctgaatcagg	atattcttct	aatacctgga	atgctgtttt	cccggggatc	4260
gcagtgggtga	gtaacctatgc	atcatcagga	gtacggataa	aatgcttgat	ggtcgggaaga	4320
ggcataaatt	ccgtcagcca	gtttagtctg	accatctcat	ctgtaacatc	attggcaacg	4380
ctacctttgc	catgtttcag	aaacaactct	ggcgcatcgg	gcttcccata	caatcgatag	4440
attgtcgcac	ctgattgccc	gacattatcg	cgagcccatt	tatacccata	taaatcagca	4500
tccatgttgg	aatttaaatcg	cggcctcgag	caagacgttt	cccgttgaat	atggctcata	4560
acacccttgg	tattactggt	tatgtaagca	gacagtttta	ttgttcatga	tgatataatt	4620
ttattcttgg	caatgtaaca	tcagagattt	tgagacacaa	cgtggccttc	ccccccccc	4680
cattattgaa	gcatttatca	gggttattgt	ctcatgagcg	gatacatatt	tgaatgtatt	4740
tagaaaaata	aacaaatagg	ggttccgcgc	acatttcccc	gaaaagtgcc	acctgacgtc	4800
taagaaacca	ttattatcat	gacattaacc	tataaaaata	ggcgtatcac	gaggcccttt	4860
cgtc						4864

<210> SEQ ID NO:16

<211> LENGTH: 4867

<212> TYPE: DNA

<213> ORGANISM: E. coli

<400> SEQ ID NO:16

tcgcgcgttt	cggatgatgac	ggtgaaaacc	tctgacacat	gcagctcccg	gagacgggtca	60
cagcttgtct	gtaagcggat	gccgggagca	gacaagcccg	tcagggcgcg	tcagcgggtg	120
ttggcgggtg	tcggggctgg	cttaactatg	cggcatcaga	gcagattgta	ctgagagtgc	180
accatatgcg	gtgtgaaata	cgcacagat	gcgtaaggag	aaaataccgc	atcagattgg	240
ctattggcca	ttgcatacgt	tgtatccata	tcataatatg	tacatttata	ttggctcatg	300
tccaacatta	ccgccatgtt	gacattgatt	attgactagt	tattaatagt	aatcaattac	360
ggggtcatta	gttcatagcc	catatatgga	gttccgcgtt	acataactta	cggtaaatgg	420
cccgcctggc	tgaccgcca	acgaccccg	cccattgacg	tcaataatga	cgtatgttcc	480
catagtaacg	ccaataggga	ctttccattg	acgtcaatgg	gtggagtatt	tacggtaaac	540
tgccacattg	gcagtcacac	aagtgtatca	tatgccaaagt	acgccccta	ttgacgtcaa	600
tgacggtaaa	tggcccgcc	ggcattatgc	ccagtacatg	accttatggg	actttcctac	660
ttggcagtac	atctacgtat	tagtcacgc	tattaccatg	gtgatgcgg	tttggcagta	720
catcaatggg	cgtggatagc	ggtttgactc	acggggattt	ccaagtctcc	acccattga	780
cgtaaatggg	agtttgtttt	ggcaccaaaa	tcaacgggac	tttccaaaat	gtcgtaaaca	840
ctccgcccc	ttgacgcaa	tgggcggtag	gcgtgtacgg	tgggaggtct	atataagcag	900
agctcgttta	gtgaaccgtc	agatcgctg	gagacgccat	ccacgctgtt	ttgacctcca	960
tagaagacac	cgggaccgat	ccagcctccg	cggcgcgaa	cgggtgcattg	gaacgcggat	1020
tccccgtgcc	aagagtgcag	taagtaccgc	ctatagactc	tataggcaca	cccctttggc	1080
tcttatgcat	gctatactgt	ttttggcttg	gggcctatac	accccgctt	ccttatgcta	1140
taggtgatgg	tatagcttag	cctatagggtg	tgggttattg	accattattg	accactcccc	1200
tattggtgac	gatactttcc	attactaatc	cataacatgg	ctctttgcca	caactatctc	1260
tattggctat	atgccaatac	tctgtccttc	agagactgac	acggactctg	tatttttaca	1320
ggatggggtc	ccatttatta	tttacaatt	cacatataca	acaacgccgt	cccccgctgc	1380
cgcagttttt	attaaacata	gcgtgggatc	tccacgcgaa	tctcgggtac	gtgttccgga	1440
catgggctct	tctccggtag	cggcgagct	tccacatccg	agccctggtc	ccatgcctcc	1500
agcggctcat	ggtcgctcgg	cagctccttg	ctcctaacag	tggaggccag	acttaggcac	1560
agcacaatgc	ccaccaccac	cagtgtgccg	cacaaggccg	tggcggtagg	gtatgtgtct	1620
gaaaaatgagc	gtggagattg	ggctcgacag	gctgacgcag	atggaagact	taaggcagcg	1680
gcagaagaag	atgcaggcag	ctgagttggt	gtattctgat	aagagtcaga	ggtaactccc	1740
gttgcgggtgc	tgtaaacggt	ggaggcgagt	gtagtctgag	cagtactcgt	tgctgccgcg	1800
cgcgcacca	gacataatag	ctgacagact	aacagactgt	tcctttccat	gggtcttttc	1860
tgcagtcacc	gtccttagat	ctgctgtgcc	ttctagtgtc	cagccatctg	ttgtttgccc	1920
ctcccccggtg	ccttccttga	ccctggaagg	tgccactccc	actgtccttt	cctaataaaa	1980
tgaggaaatt	gcacgcatt	gtctgagtag	gtgtcattct	attctggggg	gtggggtggg	2040
gcaggacagc	aagggggagg	attgggaaga	caatagcagg	catgctgggg	atgcgggtggg	2100
ctctatggcc	gctgcggcca	ggtgctgaag	aattgacccg	gttcctcctg	ggccagaaag	2160
aagcaggcac	atccccctct	ctgtgacaca	ccctgtccac	gccccgtggt	cttagttcca	2220

```

gccccactca taggacactc atagctcagg aggggtccgc cttcaatccc acccgctaaa 2280
gtacttggag cgggtctctcc ctccctcatc agcccaccaa accaaacctc gcctccaaga 2340
gtgggaagaa attaaagcaa gataggctat taagtgcaga gggagagaaa atgcctccaa 2400
catgtgagga agtaatgaga gaaatcatag aatttcttcc gcttcctcgc tcaactgactc 2460
gctgcgcctcg gtcgttcggc tgcggcgagc ggtatcagct cactcaaagg cggtaatagc 2520
gttatccaca gaatcagggg ataacgcagg aaagaacatg tgagcaaaaag gccagcaaaa 2580
ggccagggaac cgtaaaaagg ccgcgttgct ggcgtttttc cataggctcc gccccctga 2640
cgagcatcac aaaaatcgac gctcaagtca gaggtggcga aacccgacag gactataaag 2700
ataccaggcg tttccccctg gaagctccct cgtgcgctct cctgttccga ccctgccgct 2760
taccggatac ctgtccgctt ttctcccttc tgcgtccaag ctgggctgtg tgcacgaacc 2820
ctgtagggtat ctcaagtcgg tgtaggtcgt tgcgtccaag ctgggctgtg tgcacgaacc 2880
ccccgttcag cccgaccgct gcgccttctc cggttaactat cgtcttgagt ccaaccgggt 2940
aagacacgac ttatcgccac tggcagcagc cactggtaac aggattagca gagcgaggta 3000
tgtaggcggg gctacagagt tcttgaagtg gtggcctaac tacggctaca ctagaagaac 3060
agtatttggg atctgcgctc tgctgaagcc agttaccttc ggaaaaagag ttggtagctc 3120
ttgatccggc aaacaaacca ccgctggtag cgggtggtttt tttgtttgca agcagcagat 3180
tacgcgcaga aaaaaaggat ctcaagaaga tcctttgatc ttttctacgg ggtctgacgc 3240
tcagtggaaac gaaaactcac gttaagggat tttggctcat agattatcaa aaaggatctt 3300
cacctagatc cttttaaatt aaaaatgaag ttttaaatac atctaaagta tatatgagta 3360
aacttgggtc gacagttacc aatgcttaat cagtgaaggc cctatctcag cgatctgtct 3420
atttcgttca tccatagttg cctgactcgg gggggggggg cgctgaggtc tgctcgtga 3480
agaaggtggt gctgactcat accaggcctg aatcgcccca tcatccagcc agaaagtga 3540
ggagccacgg ttgatgagag ctttgttgta ggtggaccag ttggtgattt tgaactttt 3600
ctttgccacg gaacgggtctg cgttgctcgg aagatgcgtg atctgaccc tcaactcagc 3660
aaaagttcga tttattcaac aaagccgccc tcccgtaag tcagcgtaat gctctgccag 3720
tgttacaacc aattaaccaa ttctgattag aaaaactcat cgagcatcaa atgaaactgc 3780
aatttattca tatcaggatt atcaatacca ttttttgaa aaagccggtt ctgtaatgaa 3840
ggagaaaact caccgaggca gttccatagg atggcaagat cctgggtatc gtctgcgatt 3900
ccgactcgct caacatcaat acaacctatt aatttcccct cgtcaaaaat aagggtatca 3960
agtgcagaaat caccatgagt gacgactgaa tccggtgaga atggcaaaa cttatgcatt 4020
tctttccaga cttgttcaac aggcagccca ttacgctcgt catcaaaatc actcgcatca 4080
accaaaccgt tattcattcg tgattgcgcc tgagcgagac gaaatacgcg atcgctgtta 4140
aaaggacaat tacaaacagg aatcgaatgc aaccggcgca ggaacactgc cagcgcatca 4200
acaatatatt cacctgaatc aggatattct tctaatacct ggaatgctgt tttcccgggg 4260
atcgagtggt tgagtaacca tgcacatca ggagtacgga taaaatgctt gatggtcgga 4320
agaggcataa attccgtcag ccagtttagt ctgacctct catctgtaac atcattggca 4380
acgctacett tgccatgttt cagaaacaac tctggcgcat cgggcttccc atacaatca 4440
tagattgtcg cacctgattg cccgacatta tccgagccc atttataccc atataaatca 4500
gcatccatgt tggaaattta tgcgggctc gagcaagacg tttcccggtg aatatggctc 4560
ataacacccc ttgtattact gtttatgtaa gcagacagtt ttattgttca tgatgatata 4620
tttttatctt gtgcaatgta acatcagaga ttttgagaca caacgtggct tcccccccc 4680
ccccattatt gaagcattta tcagggttat tgtctcatga gcggatacat atttgaatgt 4740
atttagaaaa ataaacaaat aggggttccg cgcacatttc cccgaaaagt gccacctgac 4800
gtctaagaaa ccattattat catgacatta acctataaaa ataggcgat caccaggccc 4860
tttcgctc

```

<210> SEQ ID NO:17

<211> LENGTH: 78

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: oligonucleotide

<400> SEQ ID NO:17

```

gatcaccatg gatgcaatga agagagggct ctgctgtgtg ctgctgctgt gtggagcagt 60
cttcgtttcg cccagcga

```

78

<210> SEQ ID NO:18

<211> LENGTH: 78

<212> TYPE: DNA
<213> ORGANISM:Artificial Sequence

<220> FEATURE:
<223> OTHER INFORMATION: oligonucleotide

<400> SEQ ID NO:18
gatctcgctg ggcgaaacga agactgctcc acacagcagc agcâcacagc agagccctct 60
cttcattgca tccatggt 78

<210> SEQ ID NO:19
<211> LENGTH: 27
<212> TYPE: PRT
<213> ORGANISM:Homo sapien

<400> SEQ ID NO:19
Met Asp Ala Met Lys Arg Gly Leu Cys Cys Val Leu Leu Leu Cys Gly
1 5 10 15
Ala Val Phe Val Ser Pro Ser Glu Ile Ser Ser
20 25

<210> SEQ ID NO:20
<211> LENGTH: 33
<212> TYPE: DNA
<213> ORGANISM:Artificial Sequence

<220> FEATURE:
<223> OTHER INFORMATION: oligonucleotide

<400> SEQ ID NO:20
ggtacaaata ttggctattg gccattgcat acg 33

<210> SEQ ID NO:21
<211> LENGTH: 36
<212> TYPE: DNA
<213> ORGANISM:Artificial Sequence

<220> FEATURE:
<223> OTHER INFORMATION: oligonucleotide

<400> SEQ ID NO:21
ccacatctcg aggaaccggg tcaattcttc agcacc 36

<210> SEQ ID NO:22
<211> LENGTH: 38
<212> TYPE: DNA
<213> ORGANISM:Artificial Sequence

<220> FEATURE:
<223> OTHER INFORMATION: oligonucleotide

<400> SEQ ID NO:22
ggtacagata tcgaaagcc acgttgtgtc tcaaaatc 38

<210> SEQ ID NO:23
<211> LENGTH: 36
<212> TYPE: DNA
<213> ORGANISM:Artificial Sequence

<220> FEATURE:
<223> OTHER INFORMATION: oligonucleotide

<400> SEQ ID NO:23
cacatggatc cgtaatgctc tgccagtgtt acaacc 36

<210> SEQ ID NO:24
<211> LENGTH: 39
<212> TYPE: DNA
<213> ORGANISM:Artificial Sequence

<220> FEATURE:
<223> OTHER INFORMATION: oligonucleotide

<400> SEQ ID NO:24
ggtacatgat cacgtagaaa agatcaaagg atcttcttg 39

<210> SEQ ID NO:25
<211> LENGTH: 35
<212> TYPE: DNA
<213> ORGANISM:Artificial Sequence

<220> FEATURE:
<223> OTHER INFORMATION: oligonucleotide

<400> SEQ ID NO:25
ccacatgtcg acccgtaaaa aggccgcgtt gctgg 35

<210> SEQ ID NO:26
<211> LENGTH: 4864
<212> TYPE: DNA
<213> ORGANISM:E. coli

<400> SEQ ID NO:26
tcgcgcgttt cgggtgatgac ggtgaaaacc tctgacacat gcagctcccg gagacgggtca 60
cagcttgtct gtaagcggat gccgggagca gacaagcccg tcagggcgcg tcagcgggtg 120
ttggcgggtg tcggggctgg cttaactatg cggcatcaga gcagattgta ctgagagtgc 180
accatatgcy gtgtgaaata ccgcacagat gcgtaaggag aaaataccgc atcagattgg 240
ctattggcca ttgcatacgt tgtatccata tcataatatg tacatttata ttggctcatg 300
tccaacatta ccgccatgtt gacattgatt attgactagt tattaatagt aatcaattac 360
ggggtcatta gttcatagcc catatatgga gttccgcgtt acataactta cggtaaatgg 420
ccgcctggc tgaccgccc aagacccccc cccattgacg tcaataatga cgtatgttcc 480
catagtaacg ccaataggga ctttccattg acgtcaatgg gtggagtatt tacggtaaac 540
tgcccacttg gcagtacatc aagtgtatca tatgccaaat acgcccccta ttgacgtcaa 600
tgacggtaaa ttggccgcct ggcattatgc ccagtagatg accttatggg actttcctac 660
ttggcagtac atctacgtat tagtcatcgc tattaccatg gtgatgcggg tttggcagta 720
catcaatggg cgtggatagc ggtttgactc acggggattt ccaagtctcc accccattga 780
cgtcaatggg agtttgtttt ggcaacaaaa tcaacgggac ttccaaaaat gtcgtaacaa 840
ctccgcccc a ttgacgcaaa tgggcggtag gcgtgtacgg tgggaggtct atataagcag 900
agctcgttta gtgaaccgct agatcgctcg gagacgccat ccacgctgtt ttgacctcca 960
tagaagacac cgggaccgat ccagcctccg cggccgggaa cgggtgcattg gaacgcggat 1020
tccccgtgcc aagagtgcag taagtaccgc ctatagagtc tataggccca ccccttggc 1080
ttcttatgca tgctatactg tttttggctt ggggtctata ccccccgct tcctcatgtt 1140
ataggtgatg gtatagctta gcctataggt gtgggttatt gaccattatt gaccactccc 1200
ctattggtga cgatactttc cactactaat ccataacatg gctctttgcc acaactctct 1260
ttattggcta tatgccaaata cactgtcctt cagagactga cacggactct gtatttttac 1320
aggatgggt ctcattttatt atttacaat tcacatatat aacaccaccg tccccagtgc 1380
ccgcagtttt tattaaacat aacgtgggat ctccacgcga atctcgggta cgtgttccgg 1440
acatgggctc ttctccggtg gcggcgagc ttctacatcc gagccctgct cccatgcctc 1500
cagcgactca tggctcgctcg gcagctcctt gctcctaaca gtggaggcca gacttaggca 1560
cagcacgatg cccaccacca ccagtgtgcc gcacaaggcc gtggcggtag ggtatgtgtc 1620
tgaaaatgag ctccggggagc gggcttgac cgctgacgca tttggaagac ttaaggcagc 1680
ggcagaagaa gatgcaggca gctgagttgt tgtgttctga taagagtcag aggttaactcc 1740

cggtgagggtg	ctgttaacgg	tggagggcag	tgtagtctga	gcagtactcg	ttgctgccgc	1800
gcgcgccacc	agacataata	gctgacagac	taacagactg	ttcctttcca	tgggtctttt	1860
ctgcagtcac	cgtccttaga	tctgctgtgc	cttctagttg	ccagccatct	gttgtttgcc	1920
cctcccccg	gccttccttg	accctggaag	gtgccactcc	cactgtcctt	tcctaataaa	1980
atgaggaaat	tgcatcgcac	tgtctgagta	gggtgcattc	tattctgggg	gggtgggtgg	2040
ggcagcacag	caagggggag	gattgggaag	acaatagcag	gcattgctgg	gatgcgggtg	2100
gctctatggg	tacccagggtg	ctgaagaatt	gaccgcgttc	ctcctggggc	agaaagaagc	2160
aggcacatcc	ccttctctgt	gacacaccc	gtccacgccc	ctgggtctta	gttccagccc	2220
cactcatagg	acactcatag	ctcaggagg	ctccgccttc	aatcccaccc	gctaaagtac	2280
ttggagcggt	ctctccctcc	ctcatcagcc	caccaaacca	aacctagcct	ccaagagtgg	2340
gaagaaatta	aagcaagata	ggctattaag	tgcagaggga	gagaaaatgc	ctccaacatg	2400
tgaggaagta	atgagagaaa	tcatagaatt	tcttccgctt	cctcgctcac	tgactcgctg	2460
cgctcggtcg	ttcggtctgc	gcgagcggtg	tcagctcact	caaaggcggt	aatacgggtt	2520
tccacagaat	caggggataa	cgcaaggaa	aacatgtgag	caaaaggcca	gcaaaaggcc	2580
aggaaccgta	aaaaggccgc	gttgctggcg	tttttccata	ggctccgccc	ccctgacgag	2640
catcacaaaa	atcgacgctc	aagtcagagg	tggcgaaacc	cgacaggact	ataaagatac	2700
caggcggttc	cccttggaag	ctccctcggt	cgctctcctg	ttccgaccct	gccgcttacc	2760
ggatacctgt	ccgcctttct	cccttcggga	agcgtggcgc	tttctcaatg	ctcacgctgt	2820
aggtatctca	gttcggtgta	ggtcggtcgc	tccaagctgg	gctgtgtgca	cgaaccccc	2880
gttcagcccc	accgtcgcc	cttatccggt	agatatcgte	ttgagtccaa	cccggttaaga	2940
cacgacttat	cgccactggc	agcagccact	ggtaacagga	ttagcagagc	gaggtatgta	3000
ggcggtgcta	cagagtctct	gaagtgggtg	cctaactacg	gctacactag	aaggacagta	3060
tttggtatct	gcgctctgct	gaagccagtt	accttcggaa	aaagagtgtg	tagctcttga	3120
tccggcaaac	aaaccaccgc	tggtagcggt	ggtttttttg	tttgcaagca	gcagattacg	3180
cgagaaaaaa	aaggatctca	agaagatcct	ttgatctttt	ctacggggtc	tgacgctcag	3240
tggaacgaaa	actcacgtta	agggattttg	gtcatgagat	tatcaaaaag	gatcttcacc	3300
tagatccttt	taaattaaaa	atgaagtttt	aaatcaatct	aaagtatata	tgagtaaaact	3360
tggtctgaca	gttaccaatg	cttaatcagt	gaggcactta	tctcagcgat	ctgtctattt	3420
cgttcatcca	tagttgcctg	actccggggg	ggggggggcg	tgaggtctgc	ctcgtgaaga	3480
aggtgttgct	gactcatacc	aggcctgaat	cgccccatca	tccagccaga	aagtgaggga	3540
gccacgggtg	atgagagctt	tgttgtaggt	ggaccagtgt	gtgattttga	acttttgctt	3600
tgcacaggaa	cgtctcgctg	tgtcgggaag	atgctgtgat	tgatccttca	actcagcaaa	3660
agttcgatgt	attcaacaaa	gccgcccgtc	cgtaagtca	gcgtaatgct	ctgccagtgt	3720
tacaaccaat	taaccaattc	tgattagaaa	aactcatcga	gcatacaaat	aaactgcaat	3780
ttattcatat	caggattatc	aataccatat	ttttgaaaaa	gccgtttctg	taatgaagga	3840
gaaaactcac	cgaggcagtt	ccataggatg	gcaagatcct	ggatcgggtc	tgcgattccg	3900
actcgcccaa	catcaataca	acctattaat	ttcccctcgt	caaaaataag	gttatcaagt	3960
gagaaatcac	catgagtgc	gactgaatcc	ggtgagaatg	gcaaaagctt	atgcatttct	4020
ttccagactt	gttcaacagg	ccagccatta	cgctcgctat	caaaatcact	cgcatcaacc	4080
aaaccggtat	tcattcgtga	ttgcgcctga	gcgagacgaa	atacgcgatc	gctgttaaaa	4140
ggacaattac	aaacaggaat	cgaatgcaac	cggcgcagga	acactgccag	cgcatcaaca	4200
atattttcac	ctgaatcagg	atattcttct	aatacctgga	atgctgtttt	ccccgggatc	4260
gcagtgggtg	gtaaccatgc	atcatcagga	gtacggataa	aatgcttgat	ggtcggaaga	4320
ggcataaatt	ccgtcagcca	gtttagtctg	accatctcat	ctgtaacatc	attggcaacg	4380
ctacctttgc	catgtttcag	aaacaactct	ggcgcatcgg	gcttcccata	caatcgatag	4440
attgtcgcac	ctgattgccc	gacattatcg	cgagcccat	tatacccata	taaatacagca	4500
tccatgttgg	aatttaaatc	cgccctcgag	caagacgttt	cccgttgaat	atggctcata	4560
acaccctttg	tattactggt	tatgtaagca	gacagtttta	ttgttcatga	tgatatattt	4620
ttatcttgtg	caatgtaaca	tcagagattt	tgagacacaa	cgtggctttc	cccccccccc	4680
cattattgaa	gcatttatca	gggttattgt	ctcatgagcg	gatacatatt	tgaatgtatt	4740
tagaaaaata	aacaaatagg	ggttccgcgc	acatttcccc	gaaaagtgcc	acctgacgtc	4800
taagaaacca	ttattatcat	gacattaacc	tataaaaaata	ggcgatatcac	gaggcccttt	4860
cgtc						4864

<210> SEQ ID NO:27

<211> LENGTH: 139

<212> TYPE: DNA

<213> ORGANISM: E. coli / HIV-1

<400> SEQ ID NO:27

catgggtctt ttctgcagtc accgtccttg agatctgcca ccatgggagg caagtgggtc

60

aagaggtccg tgccccaccc cgagtactac aaggactgct aaagcccggg cagatctgct 120
gtgccttcta gttgccagc 139

<210> SEQ ID NO:28
<211> LENGTH: 139
<212> TYPE: DNA
<213> ORGANISM:E. coli / HIV-1

<400> SEQ ID NO:28
catgggtcctt ttctgcagtc accgtccttg agatctgccca ccatggccgg caagtgggtcc 60
aagaggtccg tgccccaccc cgagtactac aaggactgct aaagcccggg cagatctgct 120
gtgccttcta gttgccagc 139

<210> SEQ ID NO:29
<211> LENGTH: 203
<212> TYPE: DNA
<213> ORGANISM:E. coli / HIV-1

<400> SEQ ID NO:29
catgggtcctt ttctgcagtc accgtcctta tatctagatc accatggatg caatgaagag 60
agggctctgc tgtgtgctgc tgctgtgtgg agcagtcctc gtttcgcca gcgagatctc 120
ctccaagagg tccgtgcccc accccgagta ctacaaggac tgctaaagcc cgggcagatc 180
tgctgtgcct tctagttgcc agc 203

<210> SEQ ID NO:30
<211> LENGTH: 651
<212> TYPE: DNA
<213> ORGANISM:Human Immunodeficiency Virus - 1

<400> SEQ ID NO:30
atgggtggca agtgggtcaaa acgtagtgtg cctggatggt ctactgtaag ggaaagaatg 60
agacgagctg agccagcagc agataggggtg agacgaactg agccagcagc agtaggggtg 120
ggagcagtat ctcgagacct ggaaaaacat ggagcaatca caagtagcaa tacagcagct 180
accaatgctg attgtgcctg gctagaagca caagaggatg aggaagtggg ttttcagtc 240
agacctcagg tacctttaag accaatgact tacaaggag ctgtagatct tagccacttt 300
ttaaaagaaa aggggggact ggaagggcta attcactcac agaaaagaca agatattcctt 360
gatctgtggg tctaccacac acaaggctac ttccctgatt ggcagaacta cacaccaggg 420
ccaggaatca gatttccatt gacctttgga tgggtgcttca agctagtacc agttgagcca 480
gaaaaggtag aagaggccaa tgaaggagag aacaactgct tgttacaccc tatgagccag 540
catgggatag aggaccggga gaagggaagtg ttagagtgga ggtttgacag caagctagca 600
tttcatcacg tggcccagaga gctgcatccg gagtactaca aggactgctg a 651